

A Biochemical and Structural Characterization of
Drosophila Neuroglial

Thesis by
Andrew Henry Huber

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1994

(Defended May 31, 1994)

© 1994

Andrew Henry Huber

All Rights Reserved

Dedicated to my wife, Rachel,
and my parents

Acknowledgments

My decision to apply for admission to Caltech's graduate program in Biology was influenced by several factors: Caltech's reputation for great science, no application fee (how could you lose?), proximity to my father, and the weather (which is great if you don't breathe, or so I'm told). However, by the time I was called for an interview, I had decided to study protein structure and Caltech did not seem (based on the literature I had been sent by the Division of Biology) to be a very good place to do that. I came for an interview anyway and heard a somewhat disheveled (her new car had just been repossessed) new professor give a mini-seminar on the crystal structure of the class I major histocompatibility complex molecule HLA-A2. She didn't really have a lab yet (just an office and a very mellow dog), but the work sounded great and she seemed very personable.

My decision to come to Caltech was a difficult one. There were several things I liked about the institute and the graduate program in the Division of Biology. However, I was only really interested in Pamela's lab and my conversations with her had been relatively brief. Attending a graduate school to specifically work with a new professor I didn't really know seemed more than a little risky. In retrospect, Caltech was a great choice.

I would like to thank Pamela Bjorkman, my thesis advisor, for providing an environment conducive to learning and for being supportive and patient. Patience is definitely a virtue, and although she frequently doesn't seem to be very patient, she is almost infinitely so when it really matters.

Crystallography is a rather long and complicated process and there are several people who have helped me along the way: Michael Blum, Barbara Hsu, Michael Stowell, and Alfonso Mondragón provided invaluable help with data collection and processing; Art Chirino, Bill Weis and Wim Burmeister were of great assistance during model building and

refinement; and Doug Rees, Roland Strong, Debbie Woo and Leemor Joshua-Tor provided helpful discussions and general assistance.

I would like to thank the past and present (famous and infamous) members of the Bjorkman lab for providing interesting discussion and for their help in dealing with day-to-day life in the lab. The cast of share croppers includes (in order of appearance), Greg (aka Sequi, Betty, and Barbie) Valente, Steve Clark, Eve (the incredible vanishing woman) Zoebisch, Wen-Ching Wang, Louis Gastinel, Roland Strong (definitely one of the infamous), Peggy (“Andy, you’re wrong”) Fahnestock, Dan Vaughn, Tippy Harlow, David Penny, Ilana Tamir, George Liu, Malini Raghavan, Les Weiner, Wim Burmeister, and Jennifer Johnson. The lab squatters, who have definitely made lab life more interesting, were played by Renny Feldman, Marcus Chen and Suchka.

There are several people who have provided friendship and have generally made life here more enjoyable, including the ultimate crowd (John Bowman, Eric Arn, Chris Schoenherr, Jamie Mazer, Mike Lewicki, Dan Vaughn, Paul Garrity, and Jim Knierim), Laura Brockman, Michael and Jian-Ping Stowell, and Andy and Alex Golden.

I thank my dad, Bill, and my stepmother, Diana, for providing love, a great nearby hideaway, camaraderie, good food, and a laundry machine. I would also like to acknowledge the emotional and financial support provided by my mother, Shirley, and the positive influence my stepfather, Art, had on my life. He is gone but not forgotten.

Finally, I would like to thank my wife, Wai-Sze Rachel Huber who has endured my long work hours and vicariously experienced the frustration and brief moments of satisfaction that seem to be intrinsic to graduate student life. Her love, companionship and support have had a profound effect on my life and are very important to me.

Abstract

A number of cell-cell interactions in the nervous system are mediated by immunoglobulin gene superfamily members. For example, neuroglian, a homophilic neural cell adhesion molecule in *Drosophila*, has an extracellular portion comprising six C-2 type immunoglobulin-like domains followed by five fibronectin type III (FnIII) repeats. Neuroglian shares this domain organization and significant sequence identity with L1, a murine neural adhesion molecule that could be a functional homologue. Here I report the crystal structure of a proteolytic fragment containing the first two FnIII repeats of neuroglian (NgFn1,2) at 2.0Å. The interpretation of photomicrographs of rotary shadowed Ng, the entire extracellular portion of neuroglian, and NgFn1-5, the five neuroglian FnIII domains, is also discussed.

The structure of NgFn1,2 consists of two roughly cylindrical β -barrel structural motifs arranged in a head-to-tail fashion with the domains meeting at an angle of $\sim 120^\circ$, as defined by the cylinder axes. The folding topology of each domain is identical to that previously observed for single FnIII domains from tenascin and fibronectin. The domains of NgFn1,2 are related by an approximate two fold screw axis that is nearly parallel to the longest dimension of the fragment. Assuming this relative orientation is a general property of tandem FnIII repeats, the multiple tandem FnIII domains in neuroglian and other proteins are modeled as thin straight rods with two domain zig-zag repeats. When combined with the dimensions of pairs of tandem immunoglobulin-like domains from CD4 and CD2, this model suggests that neuroglian is a long narrow molecule (20 - 30 Å in diameter) that extends up to 370Å from the cell surface.

In photomicrographs, rotary shadowed Ng and NgFn1-5 appear to be highly flexible rod-like molecules. NgFn1-5 is observed to bend in at least two positions and has a mean total length consistent with models generated from the NgFn1,2 structure. Ng molecules

have up to four bends and a mean total length of 392Å, consistent with a head-to-tail packing of neuroglian's C2-type domains.

Table of Contents

CHAPTER 1: INTRODUCTION.....	1
Selectins.....	3
Cadherins.....	6
Integrins.....	8
Immunoglobulin Gene Superfamily.....	15
Biochemical and Structural Characterization of <i>Drosophila</i> Neuroglian.....	26
References.....	29
CHAPTER 2: CRYSTAL STRUCTURE OF TANDEM TYPE III FIBRONECTIN DOMAINS FROM DROSOPHILA NEUROGLIAN AT 2.0Å.....	48
Summary.....	50
Introduction.....	50
Results.....	51
Discussion.....	59
Experimental Procedures.....	61
Acknowledgments.....	63
References.....	63
CHAPTER 3: REFINEMENT OF THE STRUCTURE: A DETAILED DESCRIPTION.....	65
Introduction.....	66
Materials and Methods.....	73
Results and Discussion.....	74
References.....	115

CHAPTER 4: IDENTIFICATION OF THE METAL ION BOUND AT THE INTERFACE BETWEEN THE FIRST TWO FIBRONECTIN TYPE III REPEATS OF <i>DROSOPHILA</i> NEUROGLIAN.....	117
Introduction.....	118
Materials and Methods.....	119
Results and Discussion.....	120
References.....	128
CHAPTER 5: CHARACTERIZATION OF THE EXTRACELLULAR DOMAINS OF <i>DROSOPHILA</i> NEUROGLIAN BY ELECTRON MICROSCOPY.....	129
Introduction.....	130
Materials and Methods.....	131
Results and Discussion.....	133
References.....	142
APPENDIX A: CRYSTALLIZATION AND STOICHIOMETRY OF BINDING OF A COMPLEX BETWEEN A RAT INTESTINAL FC RECEPTOR AND FC.....	143
Summary.....	145
Introduction.....	145
Isolation and Crystallization of FcRn-Fc complexes.....	146
Demonstration that Two FcRn Molecules Can Bind to a Single Fc.....	147
Isothermal Titration Calorimetry to Measure Stoichiometry of Binding in Solution.....	148
Discussion.....	149
References.....	150
APPENDIX B: THE RELATIVE AFFINITY OF MONOCLONAL AND POLYCLONAL IGG FOR A RAT INTESTINAL FC RECEPTOR..	152

Introduction.....	153
Materials and Methods.....	156
Results and Discussion.....	158
References.....	163

List of Tables

2-1	Data collection statistics.....	52
2-2	Phasing statistics.....	52
2-3	Refinement statistics.....	53
3-1	Statistics for refinement cycles 1 to 5.....	77
3-2	Statistics for refinement cycles 6 to 21.....	89
4-1	Data collection statistics.....	125
4-2	Difference map peaks at ion binding site.....	126
4-3	Refined values for ion occupancy.....	127
5-1	Molecular measurements from photomicrographs.....	137
5-2	Number of FnIII repeats in arms of bent NgFn molecules.....	137
B-1	Estimated concentration of IgG required to inhibit FcRn binding of an ¹²⁵ I labeled rat IgG2a by 50 percent.....	159

List of Figures

1-1	A schematic view of the three known selectins.....	4
1-2	Average amino acid identity for domains from different cadherins.....	7
1-3	Structural features of integrin receptors.....	11
1-4	Schematic representation of a typical immunoglobulin structure.....	17
1-5	Schematic representation of the folding topologies of FnIII and C2-type and V-type immunoglobulin domains.....	19
1-6	A schematic view of L1, F11 and N-CAM.....	22
2-0	Neuron cover.....	49
2-1	SDS-polyacrylamide gel electrophoresis of glycosylated and deglycosylated forms of the chymotryptic fragment of the five Fn-III repeat protein.....	51
2-2	Initial and refined electron density maps.....	53
2-3	The structure of NgFn1,2 and sequence alignment with TnFn3.....	54
2-4	Superposition of NgFn1, NgFn2, and TnFn3.....	56
2-5	Environment around the sodium ion-binding site.....	57
2-6	Location of the pseudosymmetry axis relating NgFn1 and NgFn2.....	58
2-7	Proposed model for the alignment of Fn-III domains in series.....	59
3-1	Modified Procheck output from analysis of model produced by refinement cycle 4.....	80
3-2	Modified Procheck output from analysis of final model.....	97
3-3	Plot of Pep_flip value versus residue number.....	106
3-4	Plot of backbone real space correlation coefficient as a function of residue number.....	109
3-5	Plot of side chain real space correlation coefficient as a function of residue number.....	111
3-6	Plot of backbone temperature factor versus residue number.....	114

5-1	Measurement of segments from a bent molecule.....	132
5-2	Representative photomicrographs of rotary shadowed molecules juxtaposed with schematic models of the observed proteins.....	138
5-3	Photomicrograph images of Ng molecules with at least four bends.....	140
5-4	Schematic models of N-CAM- and neuroglial-mediated intercellular adhesion.	141
A-1	FcRn-Fc crystals.....	146
A-2	SDS/15% PAGE demonstration that washed crystals contain both Fc and FcRn proteins.....	146
A-3	Demonstration that a single Fc can bind two FcRn molecules.....	147
A-4	Calorimetric titration of FcRn and a human IgG1 monoclonal antibody.....	148
B-1	Results of competitive inhibition assays for purified subclasses of IgG.....	160
B-2	Results of polyclonal IgG competitive inhibition assays.....	162

Chapter 1

Introduction

The development of Metazoan organisms required the evolution of mechanisms for specific cell adhesion. The formation and maintenance of tissues, cell motility, proliferation and cellular differentiation all rely on the ability of cells to adhere to other cells and the extracellular matrix. This adhesion is mediated by a diverse collection of proteins, comprising homophilic and heterophilic adhesion molecules that are expressed in a temporally and spatially regulated manner. The fundamental role played by cell adhesion in tissue morphogenesis was first demonstrated with dissociated embryonic cells, which were found to sort and re-aggregate to form tissue-like structures¹⁻³. This phenomenon is now believed to be a general property of tissues from multicellular animals. These and more recent studies have led to the classification of cell adhesion molecules (CAMs) as “morphoregulatory” proteins⁴.

Cellular adhesion is not a passive process simply mediated by an adhesion molecule’s affinity for a receptor. Adhesion is an active process in which many CAMs are believed to direct changes in the cytoskeleton, a cell’s force generating machinery, through their cytoplasmic domains^{5,6}. This view is supported by experiments in which CAM cytoplasmic domains have been deleted, resulting in the cell-surface expression of stable molecules that can bind ligand but are unable to mediate cellular adhesion⁷⁻¹⁰. Some CAMs have also been demonstrated to interact with second messenger systems^{11,12}. The cytoplasmic domains of several adhesion molecules are alternatively spliced, resulting in the production of proteins with identical extracellular regions but different patterns of expression and presumably different functions^{6,11,13,14}.

Sequence analysis has revealed that several vertebrate adhesion molecules have Arthropod homologues, indicating that the progenitors of some adhesion molecule families were present before the divergence of Chordates and Arthropods¹⁵. Most CAMs belong to

one of four major families of adhesion molecules: the selectins, the cadherins, the integrins or the immunoglobulin superfamily.

Selectins

Specific intercellular adhesion plays a vital role in the regulation of regional lymphocyte traffic and the localization of white blood cells at sites of inflammation. The extravasation of leukocytes into lymphatic and inflamed tissues is preceded by a multistep process in which the leukocyte specifically adheres to the vascular endothelium¹⁶. This adhesion occurs under conditions of high shear and is characterized by an initial step in which the leukocytes are observed to roll along the endothelium at a rate that is approximately 20 fold slower than the blood flow^{17,18}. Leukocyte rolling, which is mediated by the selectin family of adhesion molecules, is necessary but not sufficient for the activation of integrins on circulating leukocytes.^{6,19} Activated leukocyte integrins mediate a tighter adhesion that leads to extravasation^{6,16,20}. Abnormal selectin expression or function can lead a variety of ailments, including recurrent bacterial infections or a number of inflammatory disorders commonly associated with excessive leukocyte recruitment¹⁹.

Selectins are glycoproteins that consist of an amino terminal carbohydrate recognition domain (CRD), followed by an epidermal growth factor (EGF)-like motif, multiple tandem repeats with sequence similarity to complement binding domains, a transmembrane domain and a cytoplasmic tail²¹. It appears as though the CRD and EGF-like domains are both required for full adhesive function²⁰. The CRD, which is related to those found in C-type (Ca^{2+} dependent) animal lectins²², binds Ca^{2+} and directly mediates the recognition and binding of specific carbohydrate ligands²³⁻²⁵. Crystallographic structure analyses of E-selectin and a C-type lectin, mannose binding protein A, strongly suggest that a bound Ca^{2+} ion plays a direct role in selectin mediated carbohydrate recognition²⁶⁻²⁹. The role of the EGF-like domain in ligand binding is still a matter of speculation.

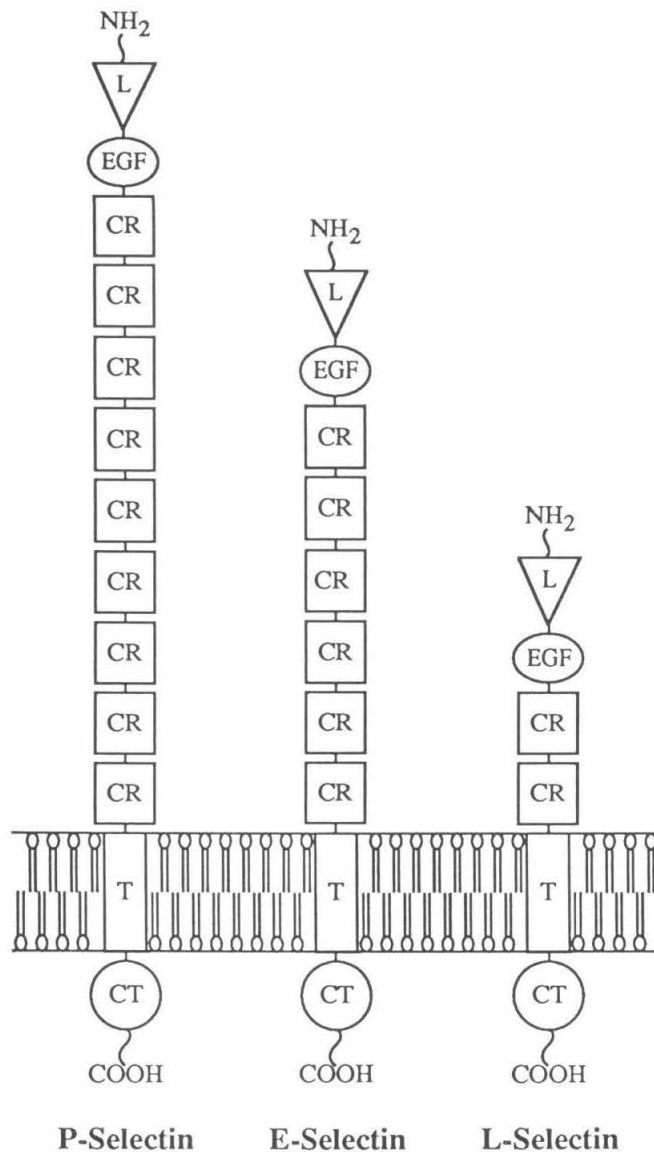


Figure 1-1 A schematic view of the three known selectins. The domain organization of each selectin was determined from primary structural information. P-selectin, E-selectin, and L-selectin are shown with carbohydrate recognition domains labeled L (lectin), EGF-like domains labeled EGF, complement regulatory-like modules labeled CR, transmembrane domains labeled T, and the cytoplasmic tail labeled CT.

There are currently three members of the selectin family¹⁹: L-selectin^{30,31} (gp90^{MEL}, LAM-1, LECAM-1), E-selectin³² (ELAM-1), and P-selectin³³ (GMP-140, PADGEM) (Figure 1-1). L-selectin is constitutively expressed on the surface of neutrophils, monocytes and subsets of lymphocytes. It has been shown to play a role in the regulation of regional lymphocyte traffic as well as in the recruitment of leukocytes at inflamed tissues. E-selectin is expressed by endothelial cells in response to a variety of inflammatory stimuli. The protein reaches the cell surface within hours of induction and binds neutrophils, monocytes, eosinophils and subsets of lymphocytes to the vascular endothelium. This effectively results in a localization of leukocytes at the inflamed tissue. P-selectin is constitutively expressed by megakaryocytes and endothelial cells, which store the protein in platelet alpha granules and Weibel-Palade bodies, respectively. P-selectin also plays a role in the localization of leukocytes at regions of inflammation. However, P-selectin is present on the surface of platelets and endothelial cells within minutes of the inducing stimulus.

E-, P-, and L-selectins have been determined to bind the sialylated, fucosylated tetrasaccharide sialyl Lewis^x (Neu5Ac α 2-3Gal β 1-4[Fuc α 1-3]GlcNAc β -) in a specific and Ca²⁺ dependent manner, suggesting that all three selectins may have similar physiological ligands^{25,34-37}. These studies have demonstrated that the proper sialic acid and fucose linkages are both important for selectin binding. While the work with small synthetic sugars has led to a greater understanding of selectin function, there is evidence that the physiological ligands may be significantly more complex. Purified P-selectin has a markedly higher affinity for myeloid HL-60 cells than CHO (Chinese hamster ovary) cells expressing sialyl Lewis^x on their surface³⁸. This difference in affinities raises the possibility that the high-affinity selectin ligands are not just glycans, but glycoproteins. High affinity glycoprotein ligands for L-selectin, P-selectin and E-selectin were recently isolated from lymphoid tissues, neutrophils and myeloid cells, respectively¹⁹. Both of the

ligands for L-selectin contain O-linked oligosaccharides that are sulfated, sialylated, and fucosylated. The ligands for all three selectins have several potential sites for O-linked oligosaccharides and only a few sites for N-linked glycans. This suggests that O-linked glycans may play an important role in selectin binding. Treating the ligands with sialidase eliminates selectin binding, demonstrating that the selectins are specifically binding the ligand oligosaccharides.

Cadherins

Cadherins are transmembrane glycoproteins of 120-140 kDa that mediate Ca^{2+} dependent homophilic adhesion³⁹. These receptors are specifically associated with adherens junctions⁴⁰⁻⁴⁵ and interact with the cytoskeleton through their cytoplasmic domain. Cadherins are differentially expressed during embryonic development and have been implicated in several morphogenic processes including selective segregation, cell condensation, cell polarization and cell differentiation⁴⁶. Inhibition of cadherin function results in a loss of epithelial shape and a stimulation of cell motility and invasiveness⁴⁷⁻⁴⁹, traits commonly associated with malignant transformation. Several undifferentiated tumors have also been found to have diminished cadherin function⁵⁰⁻⁵².

There are three major subclasses within the cadherin family: E-cadherin (epithelial cadherin, also called uvomorulin), P-cadherin (placental cadherin), and N-cadherin (neural cadherin). The expression of each protein has a unique, non-exclusive distribution. Some tissues have been found to express multiple cadherin subclasses simultaneously, raising the possibility that the subclasses can be used combinatorially to produce wide range of specific adhesive interactions⁵³. Different subclasses of cadherin from the same species of organism commonly share an overall sequence identity of approximately 50% , with the amino and carboxy terminal domains being highly conserved^{5,53}. Most of the cadherins share a common domain organization^{5,54,55}: three highly conserved amino-terminal

domains of approximately 110 amino acids each (EC1-3, Ectodomains 1-3), followed by one or two membrane proximal repeats (EC4 and 5), a transmembrane domain and a highly conserved cytoplasmic tail (Figure 1-2).

	EC1	EC2	EC3	EC4	EC5	TM	CT
Mean % identity	73.8	72.4	66.3	66.1	52.0	57.6	75.5

Figure 1-2 Average amino acid identity for domains from different cadherins. The top portion of the figure is a schematic representation of the stereotypical cadherin domain organization, with domains EC1-5 (EC1-5), a transmembrane domain (TM) and a cytoplasmic tail (CT). The mean percent amino acid identity for each domain was calculated for 16 different cadherins⁵. The following sequences were used in the comparison: human⁵⁶, bovine⁵⁷, mouse⁵⁸, chicken⁵⁹, *Xenopus*⁶⁰, and zebrafish N-cadherins⁵; chicken R-cadherin⁶¹; human⁶², bovine⁵⁷ and mouse⁶³ P-cadherins; human⁶⁴, mouse⁶⁵, and chicken⁶⁶ E-cadherins; *Xenopus* EP-cadherin⁶⁷; chicken B-cadherin⁶⁸; and mouse M-cadherin⁶⁹.

Cadherins are initially synthesized as precursor polypeptides that contain additional amino-terminal sequences (129 amino acids in E-cadherin) not found in the mature proteins. Experiments eliminating the endogenous 'presequence' cleavage site in E-cadherin demonstrated that the precursor molecule was unable to mediate adhesion until the presequence was proteolytically removed⁷⁰. These results suggest that the amino terminus plays an important role in the adhesion process. This view is supported by studies demonstrating that the binding specificity of chimeric cadherin molecules is determined by the amino terminal domain^{54,71}. Further analysis has led to the identification of specific sequence motifs within the EC1 domain that appear to play a role in homophilic adhesion. One such motif is the tripeptide HAV (Histidine-Alanine-Valine), which is conserved in the

EC1 domain of virtually all cadherins⁵. An HAV synthetic peptide has been shown to block cadherin mediated morphogenesis⁷² and the amino acids immediately flanking the HAV sequence have been shown to play an important role in the determination of cadherin specificity⁵⁴. Some EC1 sequences, including the HAV-region, share significant homology with sequences from the amino terminal domains of influenza strain A hemagglutinins⁷³. It is not yet clear whether extracellular domains other than EC1 play a role in cadherin mediated homophilic adhesion.

The cadherin cytoplasmic tail is essential for cadherin function¹⁰ and is highly conserved between cadherins as measured by sequence analysis⁵ and immunological data⁷⁴. The cytoplasmic tail is known to contain at least two functionally distinct regions that interact with cytoplasmic regulators of cellular adhesion, such as the catenins^{75,76}. Catenins are known to mediate the interaction of cadherins with the actin filament network⁷⁶ and thus regulate the adhesive function of cadherins^{55,75,77}.

Integrins

Integrins are transmembrane heterodimeric glycoproteins which mediate the adhesion of cells to extracellular matrix (ECM) proteins, such as fibronectin, laminin, vitronectin and collagen^{6,78,79}. These interactions require the presence of divalent cations (usually Ca^{2+} or Mg^{2+}) and appear to be the primary means by which cells adhere to the ECM⁶. More recently, members of the integrin family were also discovered to mediate cation dependent intercellular adhesion through the binding of Ig-like domains in cell surface members of the immunoglobulin gene superfamily⁸⁰⁻⁸². The majority of integrins bind the adhesive sequence Arg-Gly-Asp (RGD), which is recognized in the context of the presenting ligand. However, several members of this family have been shown to recognize different adhesive sequences⁶. Integrin expression is developmentally regulated with most integrins being expressed by multiple cell-types and most cells expressing multiple integrins^{6,78}. Integrins

were initially isolated from vertebrates but have also been found in *Drosophila*, demonstrating that the integrin family is evolutionarily ancient and arose before the divergence of protosomes and deuterostomes¹⁵.

Integrin heterodimers are formed by the noncovalent association of two type I membrane proteins, α and β (Figure 1-3). The proper association of α and β subunits is required for cell surface expression and ligand binding⁸³⁻⁸⁵. At this time, 14 different α and eight different β subunits have been identified in vertebrates⁶. Chemical crosslinking experiments and site directed mutagenesis have demonstrated that both subunits contribute to ligand specificity and binding^{6,11}. This initially suggested that a few different α and β subunits could be combinatorially mixed to generate a large number of unique receptors. However, since only 20 different heterodimers have actually been observed⁶, it now appears as though the number of viable combinations is relatively limited.

At the primary structural level, an α -chain consists of a large extracellular domain, a transmembrane domain and a short cytoplasmic tail. The extracellular domain contains seven tandem homologous repeats of approximately 50 amino acids that begin near the amino terminus of the protein^{6,79}. Three or four of these repeats, depending on the identity of the α -chain, contain putative cation binding sites homologous to the EF hands observed in other proteins that bind divalent cations^{6,79}. Some α -chains contain approximately 200 amino acids of additional sequence inserted between the second and third homologous repeats. This interactive or "I" domain is similar to collagen binding domains from a number of other proteins, raising the possibility that it plays a role in some integrin-collagen interactions^{6,79,86}.

β -subunits have a large amino-terminal domain followed by four cysteine rich repeats of approximately 40 amino acids, a transmembrane domain and a cytoplasmic tail^{6,79}. Overall, β -subunit cytoplasmic domains are larger than those of α -subunits but are still

relatively short (<50 amino acids). One notable exception is the cytoplasmic tail of β_4 , which contains over 1000 residues⁸⁷. A variety of experimental techniques, including chemical cross-linking and β -chain mutagenesis, have provided solid evidence that the β amino terminal domain is involved in heterodimer formation and ligand binding^{6,88}. The role played by the cysteine rich repeats is still unclear. Truncation experiments have recently demonstrated that these repeats are not necessary for the assembly, stability or functional activity of the α_{IIb} - β_3 integrin⁸⁸.

Unfortunately, the three-dimensional atomic structures of the integrin subunits and the nature of the dimer interface are still unknown. Some insight has been provided by electron micrographs of rotary shadowed heterodimers⁸⁸⁻⁹⁰. Each subunit is observed to form a long stalk that extends 12-15 nm from the plasma membrane before it terminates in a large globular head region formed by the association of the subunits (Figure 1-3).

Although integrin cytoplasmic domains are usually quite short, they are known to mediate integrin-cytoskeletal interactions and may be involved in the initiation of a variety of other intracellular processes. The colocalization of integrins and cytoskeletal structures is well documented and there is some biochemical evidence of direct interaction between integrin cytoplasmic domains and the cytoskeletal proteins talin and α -actinin⁶. Mutagenesis of integrin cytoplasmic domains has produced β -subunit mutants that fail to localize at focal contacts and α -subunit mutants that are constitutively active¹¹. These and other mutagenesis results suggest that integrin-cytoskeletal interactions are largely mediated by the β -subunit with the α -subunit playing a regulatory role¹¹.

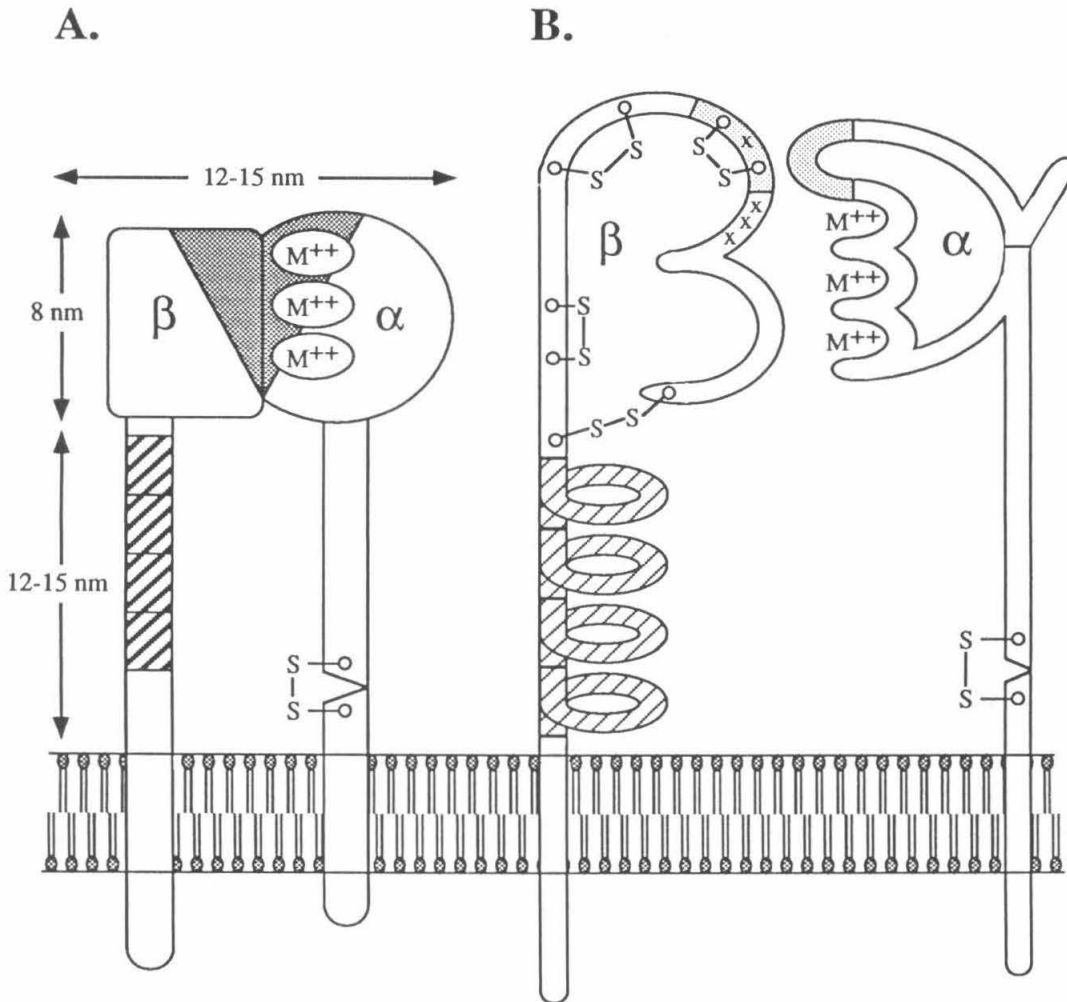


Figure 1-3 Structural features of integrin receptors. (A) shows the overall shape, as deduced from electron microscopy, as well as the putative locations of the cysteine-rich repeats of the β subunit (crosshatched) and metal binding sites in the α subunit (M^{++}). The shaded area represents the ligand-binding region that is known to be made up from both subunits based on cross-linking and binding data. (B) schematizes the arrangement of the polypeptide chains with the cysteine-rich repeats internally folded and the head region of the β subunit containing internal disulfide loops, some but not all of which are shown. A disulfide bond from the middle of the β subunit to a point close to the membrane has been proposed⁹¹ but is omitted here for clarity. X's indicate positions of mutations (of human β_2 or β_3 subunits) known to affect ligand binding or $\alpha\beta$ dimerization. The positions of alternatively spliced segments in *Drosophila* subunits are shaded. (Modified from Hynes RO. Integrins: versatility, modulation, and signaling in cell adhesion. Cell 1992;69:11-25).

Integrins are capable of initiating several intracellular signaling events, including cytoplasmic alkalinization, tyrosine phosphorylation and the elevation of cytoplasmic levels of Ca^{2+} ^{6,11}. The molecular link between ligand binding and the generation of these signals is currently unclear, but there is little doubt that the integrins are involved. Cytoplasmic alkalinization in fibroblasts, endothelial cells and lymphocytes is triggered by cellular adhesion to fibronectin or immobilized anti-integrin antibodies⁹². This change in pH is generated by activating the sodium/hydrogen ion antiporter⁹³. A similar cytoplasmic alkalinization can be triggered by phorbol ester activation of protein kinase C, suggesting that integrin initiated alkalinization might involve the activation of protein kinase C¹¹. There is also evidence that integrins can trigger tyrosine phosphorylation. Cellular adhesion to immobilized plasma fibronectin or anti-integrin antibodies results in the tyrosine phosphorylation of focal adhesion kinase (pp120, pp125^{FAK}, FAK), a protein tyrosine kinase localized in focal contacts⁹⁴⁻⁹⁶. The kinase activity of FAK is upregulated by phosphorylation, suggesting that integrin mediated adhesion initiates a tyrosine phosphorylation cascade⁹⁵. Integrin function may also be coupled to the inositol triphosphate signaling pathway. Antibody mediated aggregation of the integrin LFA-1 has been demonstrated to induce phosphoinositide hydrolysis and a transient increase in intracellular levels of Ca^{2+} ⁹⁷.

Integrins, like other CAMs, are regulated at the level of protein expression. However, it is increasingly evident that integrin mediated adhesion is also controlled by modulating the activity and specificity of existing integrin molecules. This regulation of receptor function has been observed for the β_2 integrins on leukocytes and the $\alpha_{\text{IIb}}\beta_3$ integrin on platelets. The β_2 integrins mediate several transient binding events that require the activation and deactivation of cellular adhesiveness. For example, β_2 integrins play an important role in the localization of white blood cells at sites of inflammation. These integrins are present on the surface of circulating leukocytes but are unable to bind their

ligands without the help of selectins, which mediate the initial intercellular adhesion events required for integrin activation⁶. Once activated, β_2 integrins provide the additional adhesion required for extravasation. A more specific example is provided by LFA-1 ($\alpha_L\beta_2$), which plays an important role in the adhesion of T lymphocytes to target cells and antigen presenting cells⁹⁸. In this case, crosslinking T lymphocyte antigen receptors with anti-CD3 antibodies produces a temporary increase in the affinity of LFA-1 for its ligand ICAM-1^{99,100}. A modulation of ligand affinity has also been observed for an integrin that mediates platelet aggregation, $\alpha_{IIb}\beta_3$. This integrin is present on the surface of circulating platelets, but is only capable of interacting with immobilized fibrinogen¹⁰¹. Platelet activation by thrombin, collagen, or ADP activates $\alpha_{IIb}\beta_3$, enabling it function as a receptor for soluble fibrinogen, von Willebrand factor, fibronectin, vitronectin and thrombospondin^{6,101}. There is evidence that the different observed functional states of $\alpha_{IIb}\beta_3$ correspond to different protein conformational states. This correlation has been demonstrated immunochemically and biophysically⁶.

The modulation of function observed with $\alpha_{IIb}\beta_3$ and the β_2 integrins strongly suggests that integrin activity can be controlled by intracellular signals. Thus, integrin mediated adhesion can be considered a bidirectional process in which ligand binding affects and is affected by intracellular events. Several second messenger pathways appear to be involved in the modulation of integrin function. Studies using permeabilized platelets have suggested that G proteins, inositol lipids, tyrosine kinases, protein kinase C, and Na^+/H^+ antiporters all play some role in activation, although the relative importance of these roles is yet to be determined¹¹. It is likely that some of these signaling pathways regulate the phosphorylation of integrin cytoplasmic tails, which contain serine and tyrosine phosphorylation sites¹⁰². Phosphorylation could affect integrin function by regulating *integrin-cytoskeletal* interactions.

Integrin activity and specificity can also be modulated by extracellular factors such as the identity of the divalent cations bound by the integrin¹⁰³⁻¹⁰⁶ and the integrin ligands themselves. Ligand-mediated activation has been observed with $\alpha_{IIb}\beta_3$ ¹⁰⁷. As described above, the binding of soluble fibrinogen by $\alpha_{IIb}\beta_3$ requires agonist activation of the circulating platelet. However, this “inactive” integrin can bind RGD adhesive sequences when presented as short synthetic peptides^{108,109}. The binding of RGD peptides or ligand-mimetic peptides involves the same ligand pocket as fibrinogen binding and produces a conformational change resulting in the activation of the integrin¹⁰⁷. This new conformation is immunochemically distinct from that observed after agonist activation and may represent a “high affinity” state¹⁰⁷. This suggests that there are at least three integrin conformational states with distinct ligand binding activity: an inactive conformation in which the ligand binding pocket for the adhesive sequence is relatively inaccessible, an agonist activated state with an exposed ligand binding pocket, and a high affinity state triggered by the binding of the adhesive sequence¹⁰⁷. The agonist activated state may just recognize the adhesive sequence while the high affinity state would be expected to recognize additional ligand structural elements.

*Integrins can display distinct binding specificities when expressed by different cells¹¹⁰, indicating that the regulation of integrin function is significantly more complex than a simple on-off switch. Some of this variability in function could be produced by alternative splicing. Several integrin subunits have alternatively spliced cytoplasmic tails and in some cases splicing leads to an exact cassette-type exchange of cytoplasmic domains^{6,11}. There are significantly fewer instances of alternative splicing in the putative ligand binding domains, with the only known examples being the *Drosophila* subunits PS2 (α) and PS3 (β)⁶. Additional factors, which are known to regulate integrin function and could do so in a cell specific manner, include the phospholipid composition of the surrounding plasma membrane¹¹¹ and the glycosylation state of the integrin¹¹².*

Integrins were initially divided into subfamilies by the identity of the heterodimer's β -subunit. This division was based on the premise that each β -subunit could interact with several different α -subunits, while α -subunits were limited to interaction with a single β . Several α -subunits are now known to form heterodimers with multiple β -subunits⁶, but this fact has not yet led to a change in the classification system. There are three major integrin subclasses: β_1 , β_2 and β_3 ^{78,113}. The nine members of the β_1 subfamily are expressed by a wide variety of cell types and generally mediate cell-ECM interactions, although at least one member is also involved in intercellular adhesion. These integrins are commonly referred to as VLA (very late activation) integrins, a moniker derived from the late appearance of $\alpha_1\beta_1$ (VLA1) and $\alpha_2\beta_1$ (VLA2) on T lymphocytes after *in vitro* activation. The β_2 integrins are called the leukocyte integrins or leukocyte adhesion molecules. The expression of β_2 integrins appears to be limited to leukocytes and haemopoietic precursors. The β_2 integrins, however, are not the only integrins expressed by leukocytes. The three members of this subfamily are still commonly referred to by their old names: LFA-1 ($\alpha_L\beta_2$), MAC-1 ($\alpha_M\beta_2$), and p150,95($\alpha_X\beta_2$). LFA-1 is expressed on almost all leukocytes and binds the cell surface proteins ICAM-1, ICAM-2 and ICAM-3¹¹⁴. MAC-1 and p150,95 are expressed by a large subset of leukocytes and function as receptors for inactive complement component C3b and fibrinogen. MAC-1 also appears to bind coagulation factor X and ICAM-1. The β_3 subfamily contains two members: $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$. As described above, $\alpha_{IIb}\beta_3$ is present on the surface of platelets and mediates platelet aggregation. $\alpha_V\beta_3$ has a much wider tissue distribution and acts as a receptor for a number of adhesive ligands, including fibronectin, vitronectin, fibrinogen and von Willebrand factor.

Immunoglobulin Gene Superfamily

The immunoglobulin gene superfamily (IgSF) contains a wide variety of proteins involved in cellular recognition and adhesion. All members of this superfamily contain at

least one and usually multiple domains with an immunoglobulin-like (Ig-like) fold. Sequence analysis has revealed that many immunological molecules involved in adhesion and antigen recognition contain Ig-like domains, including T-cell receptors, class I and class II major histocompatibility molecules, CD4, CD8, LFA3, ICAM-1 and Thy-1^{115,116}. Several neuronal proteins have also been determined to be IgSF members based on their deduced amino acid sequences. These proteins are thought to play a role in the development and maintenance of vertebrate and insect nervous systems. N-CAM, MAG, L1, F11, fasciclin II and neuroglian are a few of the IgSF proteins selectively expressed in the nervous system¹¹⁷. Some IgSF members, including Thy-1, CD4 and MRC OX-2, are expressed in the immune and nervous systems^{116,117}. It is not yet known whether these proteins perform similar functions in both tissue systems.

The immune and nervous systems are complex entities. In each case, the development, maintenance and operation of the system requires the formation highly specific interactions between a large number of phenotypically distinct cells. In both instances, many of these interactions are mediated by members of the immunoglobulin gene superfamily, suggesting that both systems evolved from a common evolutionary ancestor. The progenitor IgSF molecule may have been very similar to the P₀ myelin protein, a homophilic neural adhesion molecule with a single V-type (see below) extracellular domain¹¹⁵.

The immunoglobulin fold, as the name implies, was first structurally characterized in immunoglobulins, which are composed of four polypeptide chains: two identical light chains and two identical heavy chains. Disulfide bonds stabilize the light chain-heavy chain heterodimers and the heavy chain homodimer (Figure 1-4).

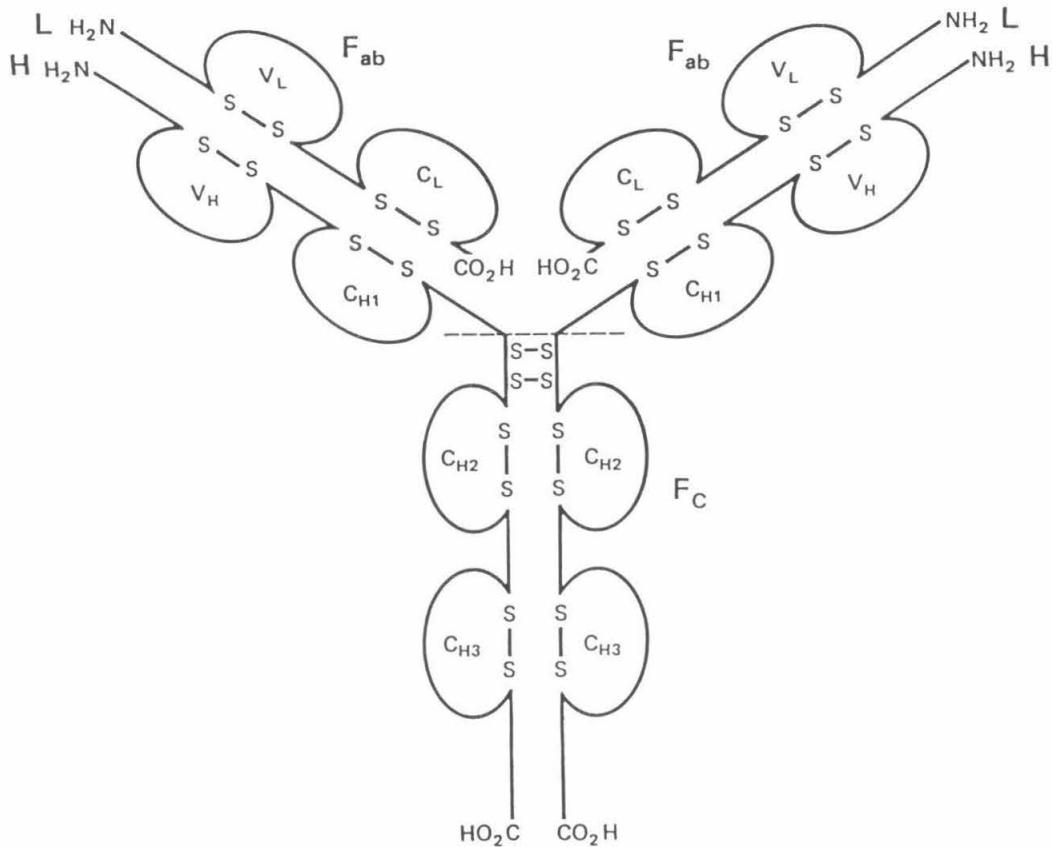


Figure 1-4 Schematic representation of a typical immunoglobulin structure.

The figure shows the L (light) and H (heavy) polypeptide chains as solid lines, with the intramolecular disulfides linking cysteine residues, about 60 residues apart in the primary structure, that are characteristic of each immunoglobulin domain. The site of cleavage by papain is shown by the dashed line; this cleavage yields two Fab fragments and one Fc. The antigen-binding sites are at the tips of the Fab arms. (From Creighton TE. *Proteins: structures and molecular properties*. (2nd ed.) New York: W. H. Freeman and Company, 1993)

Crystallographic analysis of light chain homodimers (Bence-Jones proteins) and immunoglobulin G Fab (Fragment, antigen binding) fragments defined two structurally related immunoglobulin folds: constant (C-type) domains and variable (V-type) domains^{118,119}. Immunoglobulin domains generally consist of 70-110 amino acids that form an antiparallel β -barrel with Greek key topology^{115,116,120}. The two β -sheets that form the barrel are covalently joined by a highly conserved disulfide bond. The C-type domain contains seven antiparallel β -strands that form opposing sheets of three and four strands (Figure 1-5). The four strand face is involved in the formation of C-type domain dimers. V-type domains are generally larger than C-type, with several of the “extra” residues participating in the formation of two additional antiparallel β -strands. These additional strands convert the C-type three β -strand sheet into a five strand sheet involved in the formation of V-type domain dimers (Figure 1-5). The loop connecting the additional strands forms complementarity determining region 2 (CDR2), which plays a direct role in antigen binding.

Immunoglobulin-like domains are identified at the primary structural level using several criteria: sequence similarity, secondary structure predictions, and the presence of “invariant” cysteine and tryptophane residues^{116,121}. Sequence comparisons have also led to the identification of a third type of Ig-like domain, which was named C2-type because it appeared to lack the extra β -strands characteristic of V-domains^{116,121}. V-type and C2-type domains appear to be widely distributed in proteins of the immune and nervous systems, while C-type (now C1-type) domains have only been identified in the immune system¹¹⁷. The three Ig-like folds have overlapping but distinct sequence motifs, with the

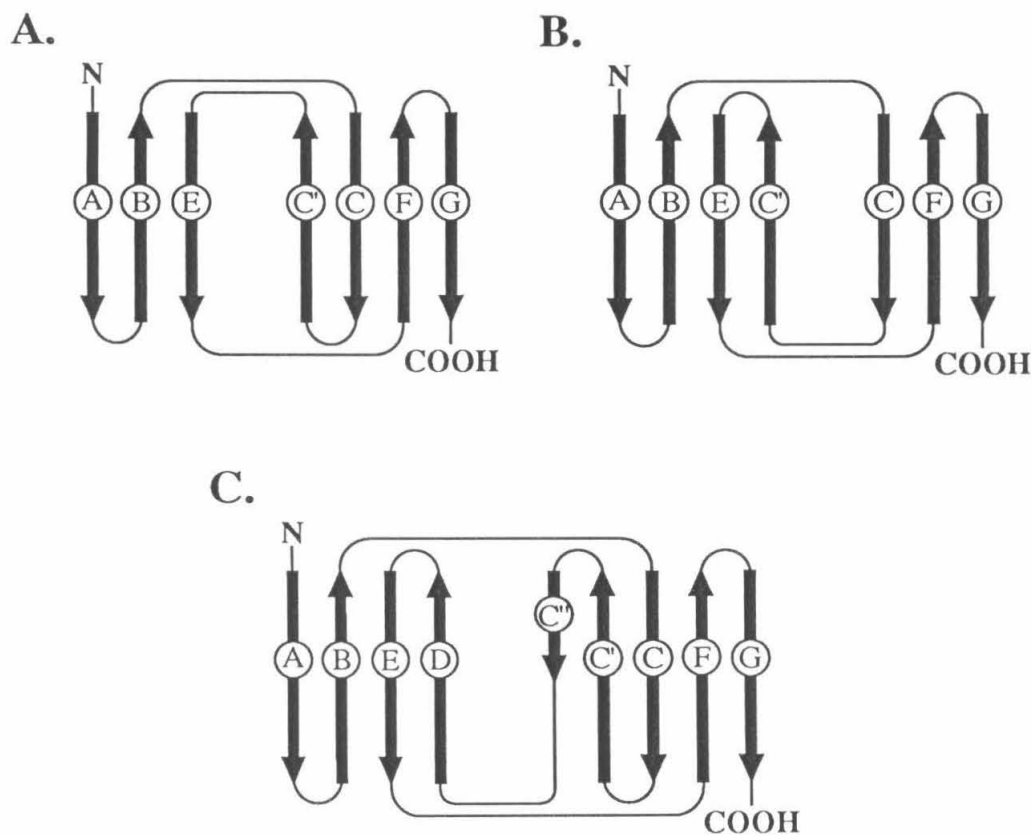


Figure 1-5 Schematic representation of the folding topologies of FnIII (A), and C2-type (B) and V-type (C) immunoglobulin domains. Each domain is composed of two β -sheets. β -strands are shown as arrows, with the direction of the arrow indicating the amino to carboxy orientation of the polypeptide chain. Turns and loops are represented as lines. For each domain, the relative orientation of the two sheets in three dimensions can be obtained by closing the topology diagram as if the sheets were pages in an open book (lying face-up).

C2-type motif being a hybrid of those originally found in C1-type and V-type domains. It has been hypothesized that this “new” Ig-like domain is actually the most evolutionarily ancient¹¹⁶. The sequences of different Ig-like folds are generally distinguished using a combination of several criteria, including the size of the domain, its predicted secondary structure, the number of residues between conserved cysteines, and the domain’s sequence similarity to known Ig-like domains.

Several members of the IgSF have been found to play important roles in the development of the nervous system. These proteins are expressed in a spatially and temporally regulated manner and are known to be involved in a variety of processes, including the adhesion of neural cells, neurite fasciculation and elongation, and axonal guidance¹¹⁷. An examination of the extracellular domains of nervous system IgSF molecules suggests that these proteins can be roughly divided into three categories: only Ig-like extracellular domains (amalgam, MAG, P₀ and Thy1), C2-type and fibronectin type III (FnIII) extracellular domains (N-CAM, L1, axonin-1, F11, fasciclin II and neuroglian), and C2-type domains plus domains other than FnIII repeats (neurocan, trk)^{12,117}.

FnIII domains were initially identified as repeats of approximately 90 amino acids in the extracellular matrix (ECM) protein fibronectin¹²². They are characterized by a conserved pattern of hydrophobic residues and have been identified in a wide range of proteins, including several ECM proteins, neural adhesion molecules, cytokine receptors, cytoplasmic muscle proteins, receptor protein kinases, receptor protein phosphatases and bacterial enzymes that cleave carbohydrate¹²³. When present, FnIII domains are commonly found as multiple tandem repeats. Fibronectin itself contains 15-17 tandem FnIII domains, with variation arising through alternative splicing¹²⁴. The amino acid sequence Arg-Gly-Asp (RGD) in the tenth fibronectin FnIII repeat (FnFn10) functions as a ligand for several integrins^{6,125}. Although the RGD tripeptide alone can support integrin

mediated adhesion, full adhesive function appears to require the presence of FnIII domains 8-10, suggesting that the integrin binding site may actually span multiple domains¹²⁶. RGD tripeptides have also been found in the FnIII domains from other proteins, such as the ECM protein tenascin¹²⁷ and the neural adhesion molecules TAG-1¹²⁸, Ng-CAM¹²⁹ and neurofascin¹³⁰. It is not yet clear, however, that these sequences play a role in adhesion. In most cases, FNIII repeats do not contain a known “adhesive sequence” and the function of these domains is still unknown. It is possible that some FnIII repeats function merely as spacers involved in the optimal placement of “active” protein domains or the linking together of different functional domains in complex proteins such as fibronectin. However, it is still too premature to assign these domains such a passive role.

The three-dimensional structures of the third tenascin FnIII repeat (TnFn3) and the tenth fibronectin FnIII repeat (FnFn10) have been determined by x-ray crystallography and NMR¹³¹⁻¹³⁴. These structures revealed that FnIII domains are β -barrels composed of two antiparallel β -sheets: a three strand sheet and a four strand sheet. The FnIII domain topology is that of a C1-type domain with strand C' having switched sheets, becoming a member of the CFG β -sheet rather than the ABE sheet as found in C1-type domains (Figure 1-5). TnFn3 and FnFn10 both contain RGD sequences in their F-G loops that can mediate integrin binding. It is likely that the structural similarities shared by FnIII and Ig-like domains are the product of convergent evolution. This view is supported by the fact that the two domains have different patterns of conserved hydrophobic residues which contribute to the formation of distinctively packed hydrophobic cores¹³¹.

Neural IgSF members with extracellular regions composed of FnIII and C2-type domains can be loosely categorized by structural homology to one of three neural glycoproteins: N-CAM, F11, or L1. These proteins contain 7-11 extracellular domains with the FnIII and C2-type domains segregated into distinct clusters of tandem repeats (Figure 1-6). Although the large number of tandem domains in these proteins may serve

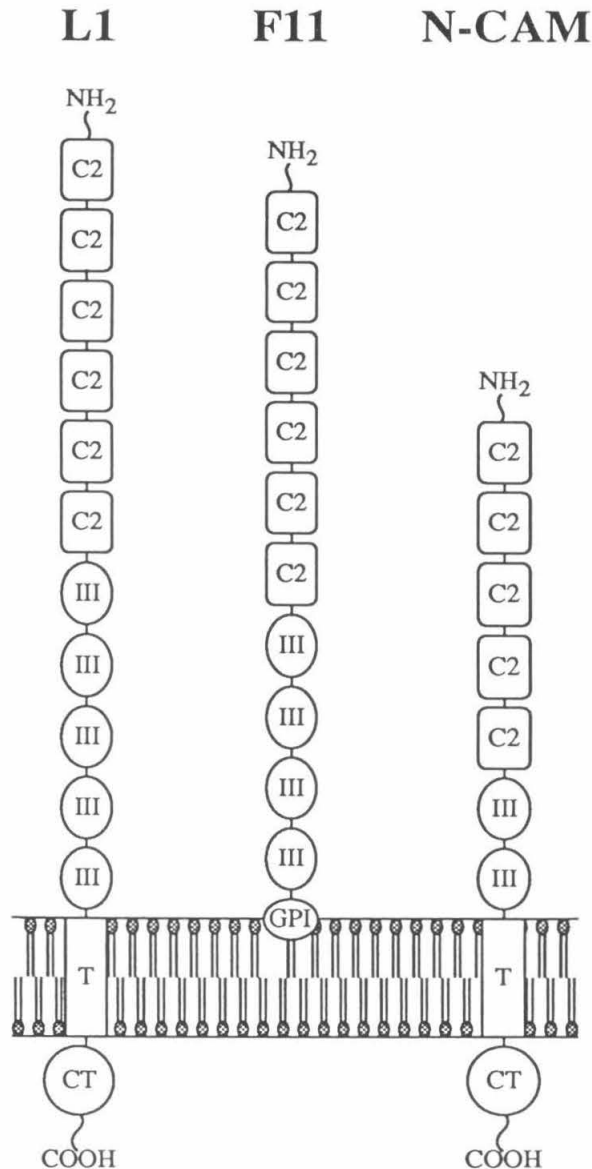


Figure 1-6 A schematic view of L1, F11 and N-CAM. Each of these proteins is representative of a subgroup of neural IgSF members containing FnIII and C2-type domains. The domain organization of each protein was determined from primary structural information. C2-type domains are labeled C2, FnIII repeats are labeled III, transmembrane domains are labeled T, cytoplasmic tails are labeled CT, and GPI (glycosylphosphatidylinositol) linkages are labeled GPI. N-CAM also exists in an alternatively spliced form (not shown) in which the transmembrane and cytoplasmic domains are absent and the protein is anchored to the lipid bilayer by glycosylphosphatidylinositol.

only to place an amino-terminal binding site far into extracellular space, an equally likely, non-exclusive alternative is that they represent multiple distinct sites for intermolecular interactions.

N-CAM was the first neural member of the IgSF to be identified and is perhaps the best characterized¹³⁵. N-CAM mRNA is subject to alternative splicing which is developmentally regulated and may produce up to 192 different isoforms of N-CAM, all of which have an extracellular region consisting of five C2-type domains followed by two FnIII repeats¹³⁶. N-CAM is expressed by a wide variety of tissues during development, including the nervous system, fetal kidney and muscle⁴. N-CAM's extracellular C2-type domains mediate Ca²⁺-independent homophilic adhesion and bind to the ECM component heparin¹³⁷. Homophilic adhesion may be regulated, at least in part, by the quantity of sialic acid attached to the protein and appears to be highly cooperative, with higher N-CAM surface density leading to a nonlinear increase in the rate of liposome aggregation.^{138,139}. There is also evidence that N-CAM interacts with cytoskeletal elements^{140,141} and secondary messenger pathways¹⁴². Although no specific ligands for the two FnIII repeats have been identified, these domains have been found to promote some functions characteristic of intact N-CAM, including the spreading of neuronal cell bodies¹⁴³. An insect structural homologue of N-CAM, Fasciclin II, was initially identified in grasshopper¹⁴⁴ and later in *Drosophila*¹⁴⁵. Fasciclin II and N-CAM share a common extracellular domain organization and 25-30% sequence identity¹⁴⁵. Although Fasciclin II and N-CAM are structurally homologous and almost certainly shared a common ancestor, their different patterns of expression suggest that they are not functionally homologous.

Members of the F11 subgroup are GPI (glycosylphosphatidylinositol) linked proteins with extracellular regions consisting of six C2-type domains followed by four FnIII repeats. This subgroup includes F11¹⁴⁶ and axonin-1¹⁴⁷ from chick as well as their

species homologues F3 (mouse)¹⁴⁸ and TAG1 (rat)¹²⁸ respectively. These proteins are expressed in lipid-linked and soluble forms. F11 is most commonly expressed as a membrane bound protein and axonin-1 is usually soluble¹². It is not yet clear why these proteins are expressed in soluble and membrane bound forms. Members of this subgroup are expressed on the surface of axons and appear to mediate axonal growth and fasciculation through heterophilic binding¹². F11 and axonin-1 have been shown to interact with the neural adhesion protein Ng-CAM^{149,150}. F11 has also been determined to bind the ECM proteins restrictin and tenascin^{149,151}. Although the proteins of this subgroup lack any direct mechanism of transducing signals across the plasma membrane, they could potentially mediate cis-heterophilic interactions (interactions with other molecules within the same membrane) with a coreceptor capable of transducing signals.

The vertebrate L1 subgroup of neural IgSF molecules includes murine L1¹⁵², rat NILE¹⁵³, chick Ng-CAM¹²⁹, chick Nr-CAM¹⁵⁴ and chick neurofascin¹³⁰. L1 and NILE are probably the same protein from different species. L1 and Ng-CAM were initially classified as species homologues but this view has recently been challenged^{155,156}. The members of the L1 subgroup have an extracellular region composed of six C2-type domains followed by five FnIII repeats, a transmembrane domain and a cytoplasmic tail of approximately 110 amino acids. The cytoplasmic tail is the most highly conserved domain shared by L1, Ng-CAM, Nr-CAM and neurofascin, suggesting that this region is important for function¹². The cytoplasmic tail might mediate interactions with cytoskeletal elements. This view is supported by the observed colocalization of Ng-CAM and actin in the filopodia of extending growth cones¹⁴⁰. There is also experimental support for L1 mediated signal transduction that may involve G-proteins^{157,158}.

L1 expression is developmentally regulated and is largely restricted to the surface of axons where it is thought to play a role in neurite extension and the formation of axonal fascicles¹¹⁷. L1, Ng-CAM and Nr-CAM appear to mediate Ca²⁺ independent homophilic

adhesion and heterophilic adhesion events^{150,156,159,160}. L1 interacts with axonin-1 and Ng-CAM has been demonstrated to interact with axonin-1 and F11. There is also evidence that L1 is involved in cis-heterotypic interactions with N-CAM^{161,162}. The heterotypic ligand for Nr-CAM is still uncharacterized. The third FnIII repeats of Ng-CAM and neurofascin contain an RGD tripeptide predicted to be in approximately the same location as that found in FnFn10 and TnFn3¹². Thus, these proteins may also interact with RGD binding integrins.

A homologue of L1 has also been identified in *Drosophila* and grasshopper^{15,163}. *Drosophila* neuroglian consists of six C2-type domains followed by five type III fibronectin repeats, a transmembrane domain and an alternatively spliced cytoplasmic tail^{14,163}. The neuroglian protein sequence is highly homologous to those of chick Ng-CAM, chick F11 and mouse L1, with neuroglian and L1 sharing greater than 28% sequence identity over the entire protein¹⁶³. This suggests that these vertebrate and insect proteins shared a common evolutionary ancestor. Tissue specific and developmentally regulated alternative splicing of neuroglian mRNA produces at least two isoforms of the protein that differ only in the size and composition of the cytoplasmic tail¹⁴. The short form of neuroglian has an 85 amino acid cytoplasmic tail and is widely expressed by neuronal and non-neuronal tissues in the developing embryo. This isoform may function as a more general adhesion molecule. The long form of neuroglian has a 148 amino acid cytoplasmic domain and a more restricted pattern of expression, appearing only on neurons of the central nervous system and a few neurons and support cells of the peripheral nervous system. The restricted expression pattern of the neuroglian long form more closely resembles the expression pattern of L1 in mice¹⁴. Neuroglian mediates Ca²⁺-independent homophilic adhesion with neuroglian expressing cells sorting away from cells expressing other *Drosophila* adhesion molecules, including fasciclin I and fasciclin II (A. Bieber et al., unpublished results). The similarities in primary structure and expression patterns shared

by neuroglian and L1, combined with neuroglian's ability to mediate homophilic adhesion, strongly suggest that neuroglian plays a role in neural adhesion and axon fasciculation.

Biochemical and Structural Characterization of *Drosophila* Neuroglian

Although biochemical characterization and mutagenesis studies have provided a wealth of information concerning the structure and function of many cellular adhesion molecules, it is often difficult to interpret these data without a three-dimensional atomic structure. At this point in time, there appears to be a remarkable shortage of known adhesion molecule structures given the large number of characterized adhesion proteins. This dearth of tertiary structural information is undoubtedly the result of several factors.

Many of the proteins discussed in the previous sections were discovered and cloned only recently. Thus, they have simply not been available for structure determinations. Now that the genes are available, however, it should be possible to produce sufficient quantities of pure protein for crystallographic or NMR structure determinations.

An incomplete characterization of molecular function is almost certainly another limiting factor. The biological role played by most adhesion molecules is still only partially understood and in many cases the protein domains responsible for the observed interactions have not yet been identified. Thus, for many adhesion proteins, there may not be enough biochemical data to entice crystallographers and NMR spectroscopists into carrying out a structure determination. Results from biochemical and mutagenesis experiments are rapidly improving this situation and have led to the identification of extracellular and intracellular ligands as well as ligand binding sites.

The sheer size and probable tertiary structure of many adhesion proteins also make them unlikely candidates for a structure determination. The extracellular portion of the

selectins, *cadherins* and most of the Ig-superfamily members contain multiple copies of a few domain motifs. These domains are thought to form linear arrays in the folded proteins, resulting in long and possibly highly flexible structures. Most of these adhesion molecules are too large to have their structures determined by NMR. While their size does not preclude a crystallographic structure determination, a high degree of flexibility might. Flexible proteins frequently produce poorly diffracting crystals, if they crystallize at all, because immobilization of the protein in a ordered crystalline lattice is a highly unfavorable process. Thus, in many cases, crystallographers and NMR spectroscopists have to resort to a “divide and conquer” approach, in which a model of the entire protein is pieced together from the structures of stable protein fragments.

Here I report a biochemical and crystallographic characterization of the *Drosophila* neural cell adhesion molecule neuroglian. As described earlier, neuroglian has an extracellular region consisting of six C2-type domains followed by five FnIII repeats. The whole extracellular portion of neuroglian and a fragment containing the five FnIII repeats were successfully crystallized, but failed to diffract to high enough resolution for a structure determination. Since these proteins were expected to be long, flexible molecules, the poor quality of the diffraction was not surprising and the divide and conquer approach was adopted. Proteolysis of the five FnIII repeat protein was found to produce a stable fragment containing the first two FnIII repeats (NgFn1,2). The functional role played by any of the neuroglian FnIII repeats is still unknown, although the first two FnIII domains from the related protein L1 have been shown to promote neurite outgrowth.

While the structures of single FnIII repeats from ECM proteins have been reported previously (see TnFn3 and FnFn10 above), these domains do not share any statistically significant sequence identity with the FnIII repeats in neural adhesion molecules. The neural domains were identified as FnIII repeats by the presence of a characteristic pattern of conserved hydrophobic residues. These neural repeats were predicted to share the same

folding topology as the ECM domains, but this had not been confirmed experimentally. When FnIII repeats are present in a protein, they are usually found in groups of multiple tandem repeats. Since the previous structure solutions were of single FnIII domains, they provided very limited insight into the packing of tandem repeats. Thus, the NgFn1,2 structure was pursued, despite the lack of any definite function, because it was felt that it would answer general structural questions concerning the topology of neural FnIII repeats and the packing of tandem FnIII domains. Chapter 2 describes the generation, purification, crystallization and structure of NgFn1,2. Chapter 3 is a detailed description of the NgFn1,2 structure refinement and the quality of the final model.

One unexpected feature of the NgFn1,2 structure was the presence of a bound metal ion at the interdomain interface. Similar metal binding sites could be present in other FnIII interdomain interfaces. These sites would presumably stabilize the relative orientations of domains involved in receptor-ligand interactions and could play a role in the formation of long rigid filaments of FnIII domains. In order to better characterize the observed metal binding site, the identity of the metal ion was determined by soaking NgFn1,2 crystals in solutions containing selected metal cations. Chapter 4 describes the results of these soaking experiments.

Assuming that the relative orientation of the domains in the NgFn1,2 structure is a general feature of tandem FnIII domains, a protein composed of several FnIII repeats is predicted to form a linear rod with a two domain zig-zag repeat. This prediction and the presence of a metal binding site at the NgFn1,2 interdomain interface raised the possibility that the five neuroglial FnIII domains (NgFn) form a long, relatively stiff, straight rod. It was not possible to test this hypothesis crystallographically because good quality crystals of NgFn or the whole extracellular portion of neuroglial (Ng) could not be obtained. It was possible, however, to better characterize the structure of NgFn and Ng by examining

electron micrographs of rotary shadowed molecules. The results of these studies are described in Chapter 5.

Appendices A and B describe work I carried out prior to initiating the neuroglial project. These experiments involved the crystallization and biochemical characterization of the neonatal rat intestinal Fc receptor (FcRn) complexed with its ligand IgG or Fc. FcRn is a noncovalent heterodimer composed of a heavy chain and a β_2 -microglobulin light chain¹⁶⁴. The FcRn heavy chain shares significant sequence identity with the heavy chains of class I MHC (major histocompatibility complex) molecules¹⁶⁵, which also form noncovalent heterodimers with β_2 -microglobulin. This primary structural homology strongly suggests that FcRn and class I MHC molecules have similar three-dimensional structures despite the fact that they bind very different biological ligands. Class I MHC molecules bind peptides of approximately nine amino acids (~ 1 kDa) while FcRn binds whole IgG molecules (~ 160 kDa). The structural mechanism by which these related proteins bind such different ligands is of general interest to those studying protein-protein recognition and binding.

References

1. Townes, P.L. & Holtfreter, J. (1955). Directed movements and selective adhesion of embryonic amphibian cells. *J. Exp. Zool.* **128**, 53-120.
2. Armstrong, P.B. (1989). Cell sorting out: the self-assembly of tissues *in vitro*. *CRC Crit. Rev. Biochem. Mol. Biol.* **24**, 119-149.
3. Albelda, S. (1991). Endothelial and epithelial cell adhesion molecules. *Am. J. Cell Mol. Biol.* **4**, 195-203.
4. Edelman, G., Cunningham, B. & Thiery, J.P. *Morphoregulatory molecules* (Wiley and Sons, New York, 1990).
5. Geiger, B. & Ayalon, O. (1992). Cadherins. *Ann. Rev. Cell Biol.* **8**, 307-332.

6. Hynes, R.O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**, 11-25.
7. Hayashi, Y., Haimovich, B., Reszka, A., Boettiger, D. & Horwitz, A. (1990). Expression and function of chicken integrin $\beta 1$ subunit and its cytoplasmic domain mutants in mouse NIH 3T3 cells. *J. Cell Biol.* **110**, 175-184.
8. Solowska, J., Guan, J.-L., Marcantonio, E.E., Trevithick, J.E., Buck, C.A. & Hynes, R.O. (1989). Expression of normal and mutant avian integrin subunits in rodent cells. *J. Cell Biol.* **109**, 853-861.
9. Kansas, G.S., Ley, K., Munro, J.M. & Tedder, T.F. (1993). Regulation of leukocyte rolling and adhesion to high endothelial venules through the cytoplasmic domain of L-selectin. *J. Exp. Med.* **177**, 833-838.
10. Nagafuchi, A. & Takeichi, M. (1988). Cell binding function of E-cadherin is regulated by the cytoplasmic domain. *EMBO J.* **7**, 3679-3684.
11. Humphries, M.J., Mould, A.P. & Tuckwell, D.S. (1993). Dynamic aspects of adhesion receptor function - integrins both twist and shout. *BioEssays* **15**, 391-397.
12. Brümmendorf, T. & Rathjen, F.G. (1993). Axonal glycoproteins with immunoglobulin- and fibronectin type III-related domains in vertebrates: structural features, binding activities, and signal transduction. *J. Neurochem.* **61**, 1207-1219.
13. Reyes, A.A., Small, S.J. & Akeson, R. (1991). At least 27 alternatively spliced forms of the neural cell adhesion molecule mRNA are expressed during rat heart development. *Mol. Cell. Biol.* **11**, 1654-1661.
14. Hortsch, M., Bieber, A.J., Patel, N.H. & Goodman, C.S. (1990). Differential splicing generates a nervous system-specific form of *Drosophila* neuroglian. *Neuron* **4**, 697-709.
15. Hortsch, M. & Goodman, C.S. (1991). Cell and substrate adhesion molecules in *Drosophila*. *Ann. Rev. Cell Biol.* **7**, 505-557.

16. Lasky, L.A. (1992). Selectins: Interpreters of cell-specific carbohydrate information during inflammation. *Science* **258**, 964-969.
17. Ley, K., Gaehtgens, P., Fennie, C., Singer, M.S., Lasky, L.A. & Rosen, S.D. (1991). Lectin-like cell adhesion molecule 1 mediates leukocyte rolling in mesenteric venules *in vivo*. *Blood* **77**, 2553-2555.
18. Atherton, A. & Born, G.V.R. (1973). Relationship between the velocity of rolling granulocytes and that of the blood flow in venules. *J. Physiol. (London)* **233**, 157-165.
19. McEver, R.P. (1994). Selectins. *Curr. Opin. Immunol.* **6**, 75-84.
20. Bevilacqua, M.P. (1993). Endothelial-leukocyte adhesion molecules. *Ann. Rev. Immunol.* **11**, 767-804.
21. Lasky, L.A. & Rosen, S.D. in *Inflammation: basic principles and clinical correlates* (eds. Gallin, J.I., Goldstein, I.M. & Snyderman, R.) (Raven Press, Ltd., New York, 1992).
22. Drickamer, K. (1988). Two distinct classes of carbohydrate recognition domains in animal lectins. *J. Biol. Chem.* **263**, 9557-9560.
23. Pigott, R., Needham, L.A., Edwards, R.M., Walker, C. & Power, C. (1991). Structural and functional studies of the endothelial activation antigen endothelial leukocyte adhesion molecule-1 using a panel of monoclonal antibodies. *J. Immunol.* **147**, 130-135.
24. Bowen, B., Fennie, C. & Lasky, L.A. (1990). The Mel-14 antibody binds to the lectin domain of the murine peripheral lymph node homing receptor. *J. Cell Biol.* **110**, 147-153.
25. Walz, G., Aruffo, A., Kolanus, W., Bevilacqua, M. & Seed, B. (1990). Recognition by ELAM-1 of the sialyl-Lex determinant on myeloid and tumor cells. *Science* **250**, 1132-1135.

26. Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W.A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* **254**, 1608-1615.
27. Weis, W.I., Drickamer, K. & Hendrickson, W.A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* **360**, 127-134.
28. Weis, W.I. (1994). Lectins on a roll: the structure of E-selectin. *Structure* **2**, 147-150.
29. Graves, B.J., al., e. & Burns, D.K. (1994). Insights into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature* **367**, 532-538.
30. Siegelman, M.H., van de Rijn, M. & Weissman, I.L. (1989). Mouse lymph node homing receptor cDNA clone encodes a glycoprotein revealing tandem interaction domains. *Science* **243**, 1165-1172.
31. Lasky, L.A., Singer, M.S., Yednock, T.A., Dowbenko, D., Fennie, C., Rodriguez, H., Nguyen, T., Stachel, S. & Rosen, S.D. (1989). Cloning of a lymphocyte homing receptor reveals a lectin domain. *Cell* **56**, 1045-1055.
32. Bevilacqua, M.P., Stengelin, S., Gimbrone, M.A., Jr. & Seed, B. (1989). Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complement regulatory proteins and lectins. *Science* **243**, 1160-1165.
33. Johnston, G.I., Cook, R.G. & McEver, R.P. (1989). Cloning of GMP-140, a granule membrane protein of platelets and endothelium: sequence similarity to proteins involved in cell adhesion and inflammation. *Cell* **56**, 1033-1044.
34. Phillips, M.L., Nudelman, E., Gaeta, F., Perez, M., Singhal, A.K., Hakomori, S. & Paulson, J.C. (1990). ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand. sialyl-Lex. *Science* **250**, 1130-1132.

35. Tiemeyer, M., Swiedler, S.J., Ishihara, M., Moreland, M., Schweingruber, H., Hirtzer, P. & Brandley, B.K. (1991). Carbohydrate ligands for endothelial-leukocyte adhesion molecule-1. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1138-1142.
36. Polley, M., Phillips, M., Wagner, E., Nudelman, A., Singhai, S., Hakomori, S. & Paulson, J. (1991). CD62 and endothelial cell adhesion molecule 1 (ELAM-1) recognize the same carbohydrate ligand: sialyl-Lewis-x. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6224-6228.
37. Foxall, C., Watson, S.R., Dowbenko, D., Fennie, C., Lasky, L.A., Kiso, M., Hasegawa, A., Asa, D. & Brandley, B.K. (1992). The 3 members of the selectin receptor family recognize a common carbohydrate epitope, the sialyl lewis oligosaccharide. *J. Cell Biol.* **117**, 895-902.
38. Zhou, Q., Moore, K.L., Smith, D.F., Varki, A., McEver, R.P. & Cummings, R.D. (1991). The selectin GMP-140 binds to sialylated, fucosylated lactosaminoglycans on both myeloid and nonmyeloid cells. *J. Cell Biol.* **115**, 557-564.
39. Nose, A., Nagafuchi, A. & Takeichi, M. (1988). Expressed recombinant cadherins mediate cell sorting in model systems. *Cell* **54**, 993-1001.
40. Boller, K., Vestweber, D. & Kemler, R. (1985). Cell-adhesion molecule uvomorulin is localized in the intermediate junctions of adult intestinal epithelial cells. *J. Cell Biol.* **100**, 327-332.
41. Geiger, B. (1989). Cytoskeleton-associated cell contacts. *Curr. Opin. Cell Biol.* **1**, 103-109.
42. Hirano, S., Nose, A., Hatta, K., Kawakami, A. & Takeichi, M. (1987). Calcium-dependent cell-cell adhesion molecules (cadherins): subclass specificities and possible involvement of actin bundles. *J. Cell Biol.* **105**, 2501-2510.
43. Volk, T., Geiger, B. & Raz, A. (1984). Motility and adhesive properties of high- and low-metastatic murine neoplastic cells. *Cancer Res.* **44**, 811-824.

44. Volk, T. & Geiger, B. (1986). A-CAM: a 135-kD receptor of intercellular adherens junctions. I. Immunoelectron microscopic localization and biochemical studies. *J. Cell Biol.* **103**, 1441-1450.
45. Volk, T. & Geiger, B. (1986). A-CAM: a 135-kD receptor of intercellular adherens junctions. II. Antibody-mediated modulation of junctional formation. *J. Cell Biol.* **103**, 1451-1464.
46. Schoenenberger, C.A., Zuk, A., Kendall, D. & Matlin, K.S. (1991). Multilayering and loss of apical polarity in MDCK cells transformed with viral K-ras. *J. Cell Biol.* **112**, 873-889.
47. Behrens, J., Mareel, M.M., VAn, R.F. & Birchmeier, W. (1989). Dissecting tumor cell invasion: epithelial cells acquire invasive properties after the loss of uvomorulin-mediated cell-cell adhesion. *J. Cell Biol.* **108**, 2435-2447.
48. Frixen, U.H., Behrens, J., Sachs, M., Eberle, G., Voss, B., Warda, A., Lochner, D. & Birchmeier, W. (1991). E-cadherin mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *J. Cell Biol.* **113**, 173-185.
49. Thiery, J.P., Boyer, B., Tucker, G., Gavrilovic, J. & Valles, A.M. (1988). Adhesion mechanisms in embryogenesis and in cancer invasion and metastasis. *Ciba Found. Symp.* **141**, 48-74.
50. Shimoyama, Y. & Hirohashi, S. (1991). Cadherin intercellular adhesion molecule in hepatocellular carcinomas: loss of E-cadherin expression in an undifferentiated carcinoma. *Cancer Lett.* **57**, 131-135.
51. Littlefield, J.W. & Whitehouse, L.L. (1990). Absence of uvomorulin in a slowly compacting variant of H6 embryonal carcinoma cells. *Somat. Cell Mol. Genet.* **16**, 191-194.
52. Hashimoto, M., Niwa, O., Nitta, Y., Takeichi, M. & Yokoro, K. (1989). Unstable expression of E-cadherin adhesion molecules in metastatic ovarian tumor cells. *Jpn. J. Cancer Res.* **80**, 459-463.

53. Takeichi, M. (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* **251**, 1451-1455.
54. Takeichi, M. (1990). Cadherins: a molecular family important in selective cell-cell adhesion. *Annu. Rev. Biochem.* **59**, 237-252.
55. Kemler, R. (1993). From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. *Trends Genet.* **9**, 317-321.
56. Walsh, F.S., Barton, C.H., Putt, W., Moore, S.E., Kelsell, D., Spurr, N. & Goodfellow, P.N. (1990). N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin gene. *J. Neurochem.* **55**, 805-812.
57. Liaw, C.W., Cannon, C., Power, M.D., Kiboneka, P.K. & Rubin, L.L. (1990). Identification and cloning of two species of cadherins in bovine endothelial cells. *EMBO J.* **9**, 2701-2708.
58. Miyatani, S., Shimamura, K., Hatta, M., Nagafuchi, A., Nose, A., Matsunaga, M., Hatta, K. & Takeichi, M. (1989). Neural cadherin: role in selective cell-cell adhesion. *Science* **245**, 631-635.
59. Hatta, K., Takagi, S., Fujisawa, H. & Takeichi, M. (1987). Spatial and temporal expression pattern of N-cadherin cell adhesion molecules correlated with morphogenetic processes of chicken embryos. *Dev. Biol.* **120**, 215-227.
60. Choi, Y. & Gumbiner, B. (1989). Expression of cell adhesion molecule E-cadherin in *Xenopus* embryos begins at gastrulation and predominates in the ectoderm. *J. Cell Biol.* **108**, 2449-2458.
61. Inuzuka, H., Miyatani, S. & Takeichi, M. (1991). R-cadherin: A novel Ca^{2+} -dependent cell-cell adhesion molecule expressed in the retina. *Neuron* **7**, 1-20.
62. Shimoyama, Y., Yoshida, T., Terada, M., Shimosato, Y., Abe, O. & Hirohashi, S. (1989). Molecular cloning of a human Ca^{2+} -dependent cell-cell adhesion molecule homologous to mouse placental cadherin: its low expression in human placental tissues. *J. Cell Biol.* **109**, 1787-1794.

63. Nose, A., Nagafuchi, A. & Takeichi, M. (1987). Isolation of placental cadherin cDNA: identification of a novel gene family of cell-cell adhesion molecules. *EMBO J.* **6**, 3655-3661.
64. Mansouri, A., Spurr, N., Goodfellow, P.N. & Kemler, R. (1988). Characterization and chromosomal localization of the gene encoding the human cell adhesion molecule uvomorulin. *Differentiation* **38**, 67-71.
65. Nagafuchi, A., Shirayoshi, Y., Okazaki, K., Yasuda, K. & Takeichi, M. (1987). Transformation of cell adhesion properties by exogenously introduced E-cadherin cDNA. *Nature* **329**, 341-343.
66. Gallin, W., Sorkin, B., Edelman, G. & Cunningham, B. (1987). Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule (L-CAM). *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2808-2812.
67. Ginsberg, D., DeSimone, D. & Geiger, B. (1991). Expression of a novel cadherin (EP-cadherin) in unfertilized eggs and early *Xenopus* embryos. *Development* **111**, 315-325.
68. Napolitano, E.W., Venstrom, K., Wheeler, E.F. & Reichardt, L.F. (1991). Molecular cloning and characterization of B-cadherin, a novel chick cadherin. *J. Cell Biol.* **113**, 893-905.
69. Donalies, M., Cramer, M., Ringwald, M. & Starzinski-Powitz, A. (1991). Expression of M-cadherin a member of the cadherin multigene family correlates with differentiation of skeletal muscle cells. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8024-8028.
70. Ozawa, M. & Kemler, R. (1990). Correct proteolytic cleavage is required for the cell adhesive function of uvomorulin. *J. Cell Biol.* **111**, 1645-1650.
71. Nose, A., Tsuj, K. & Takeichi, M. (1990). Localization of specificity determining sites in cadherin cell adhesion molecules. *Cell* **61**, 147-155.

72. Blaschuk, O.W., Sullivan, R., David, S. & Pouliot, Y. (1990). Identification of a cadherin cell adhesion recognition sequence. *Dev. Biol.* **139**, 227-229.
73. Blaschuk, O.W., Pouliot, Y. & Holland, P.C. (1990). Identification of a conserved region common to cadherins and influenza strain A hemagglutinins. *J. Mol. Biol.* **211**, 679-682.
74. Geiger, B., Volberg, T., Ginsberg, D., Bitzur, S., Sabanay, I. & Hynes, R.O. (1990). Broad spectrum pan-cadherin antibodies, reactive with the C-terminal 24 amino acid residues of N-cadherin. *J. Cell Sci.* **97**, 604-614.
75. Kintner, C. (1992). Regulation of embryonic cell adhesion by the cadherin cytoplasmic domain. *Cell* **69**, 225-236.
76. Ozawa, M., Ringwald, M. & Kemler, R. (1990). Uvomorulin-catenin complex formation is regulated by a specific domain in the cytoplasmic region of the cell adhesion molecule. *Proc. Natl. Acad. Sci. USA* **87**, 4246-4250.
77. Fujimori, T. & Takeichi, M. (1993). Disruption of epithelial cell-cell adhesion by exogenous expression of a mutated nonfunctional N-cadherin. *Mol. Cell. Biol.* **4**, 37-47.
78. Bosman, F.T. (1993). Integrins: cell adhesives and modulators of cell function. *Histochemical Journal* **25**, 469-477.
79. Reichardt, L.F. & Tomaselli, K.J. (1991). Extracellular matrix molecules and their receptors: functions in neural development. *Annu. Rev. Neurosci.* **14**, 531-70.
80. Diamond, M.S., Staunton, D.E., Marlin, S.D. & Springer, T.A. (1991). Binding of the integrin Mac-1 (CD11b/CD18) to the third immunoglobulin-like domain of ICAM-1 (CD54) and its regulation by glycosylation. *Cell* **65**, 961-971.
81. Elices, M.J., Osborn, L., Takada, Y., Crouse, C., Luhowskyj, S., Hemler, M.E. & Lobb, R.R. (1990). VCAM-1 on activated endothelium interacts with the leukocyte integrin VLA-4 at a site distinct from the VLA-4/fibronectin binding site. *Cell* **60**, 577-584.

82. Staunton, D.E., Dustin, M.L., Erickson, H.P. & Springer, T.A. (1990). The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell* **61**, 243-254.
83. Leptin, M., Bogaert, T., Lehmann, R. & Wilcox, M. (1989). The function of PS integrins during *Drosophila* embryogenesis. *Cell* **56**, 401-408.
84. Duperray, A., Troesch, A., Berthier, R., Chagnon, E., Frachet, P., Uzan, G. & Marguerie, G. (1989). *Blood* **74**, 1603-1611.
85. Kolodziej, M.A., Vilaire, G., Rifat, S., Poncz, M. & Bennett, J.S. (1991). Effect of deletion of glycoprotein IIb exon 28 on the expression of the platelet glycoprotein IIb/IIIa complex. *Blood* **78**, 2344-2353.
86. Colombatti, A. & Bonaldo, P. (1991). The superfamily of proteins with von Willebrand Factor type-A like domains: one theme common to components of extracellular matrix, haemostasis, cellular adhesion and defence mechanisms. *Blood* **77**, 2305-2315.
87. Hogervorst, F., Kuikman, I., von dem Borne, A.E.G.K. & Sonnenberg, A. (1990). Cloning and sequence analysis of β -4 cDNA: an integrin subunit that contains a unique 118 kD cytoplasmic domain. *EMBO J.* **9**, 765-770.
88. Wippler, J., Kouns, W.C., Schlaeger, E.-J., Kuhn, H., Hadvary, P. & Steiner, B. (1994). The integrin α IIb- β 3, platelet glycoprotein IIb-IIIa, can form a functionally active heterodimer complex without the cysteine-rich repeats of the β 3 subunit. *J. Biol. Chem.* **269**, 8754-8761.
89. Nermut, M.V., Green, N.M., Eason, P., Yamada, S.S. & Yamada, K.M. (1988). Electron-microscopy and structural model of human fibronectin receptor. *EMBO J.* 4093-4099.
90. Carrell, N.A., Fitzgerald, L.A., Steiner, B., Erickson, H.P. & Phillips, D.R. (1985). Structure of human platelet membrane glycoproteins IIb and IIIa as determined by electron microscopy. *J. Biol. Chem.* **260**, 1743-1749.

91. Calvete, J.J., Henschen, A. & Gonzalez-Rodriguez, J. (1991). Assignment of disulfide bonds in human platelet GPIIIa. A disulfide pattern for the beta-subunits of the integrin family. *Biochem. J.* **274**, 63-71.
92. Schwartz, M.A., Ingber, D.E., Lawrence, M., Springer, T.A. & Lechene, C. (1991). Multiple integrins share the ability to induce elevation of intracellular pH. *Exp. Cell. Res.* **195**, 533-535.
93. Schwartz, M.A., Lechene, C. & Ingber, D.E. (1991). Insoluble fibronectin activates the sodium/hydrogen ion antiporter by clustering and immobilizing integrin $\alpha 5 \beta 1$, independent of cell shape. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7849-7853.
94. Guan, J., Trevithick, J.E. & Hynes, R.O. (1991). Fibronectin/integrin interaction induces tyrosine phosphorylation of a 120 kDa protein. *Cell Regul.* **2**, 951-964.
95. Guan, J. & Shalloway, D. (1992). Regulation of focal adhesion-associated protein tyrosine kinase by both cellular adhesion and oncogenic transformation. *Nature* **358**, 690-692.
96. Kornberg, L.J., Earp, H.S., Turner, C.E., Prockop, C. & Juliano, R.L. (1991). Signal transduction by integrins: increased protein tyrosine phosphorylation caused by clustering of $\beta 1$ integrins. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8392-8396.
97. Pardi, R., Bender, J.R., Dettori, C., Giannazza, E. & Engleman, E. (1989). Heterogeneous distribution and transmembrane signaling properties of lymphocyte function-associated antigen (LFA-1) in human lymphocyte subsets. *J. Immunol.* **143**, 3157-3166.
98. Dustin, M.L. & Springer, T.A. (1991). Role of lymphocyte adhesion receptors in transient interactions and cell locomotion. *Annu. Rev. Immunol.* **9**, 27-66.
99. van Kooyk, Y., van de Weil-van Kemenade, P., Weder, P., Kuijpers, T.W. & Figdor, C.G. (1989). Enhancement of LFA-1 mediated cell adhesion by triggering through CD2 or CD3 on T lymphocytes. *Nature* **342**, 811-813.

100. Dustin, M.L. & Springer, T.A. (1989). T-cell receptor cross-linking transiently stimulates adhesiveness through LFA-1. *Nature* **341**, 619-624.
101. Phillips, D.R., Charo, I.F. & Scarborough, R.M. (1991). GPIIb-IIIa: the responsive integrin. *Cell* **65**, 359-362.
102. Tapley, P. & Rohrschneider, L. (1989). Phosphorylation of avian integrins by pp60v-src. *Adv. Protein Phosphatases* **5**, 261-278.
103. Jackson, A.M., Alexandroff, A.B., Lappin, M.B., Esuvaranathan, K., James, K. & Chisholm, G.D. (1994). Control of leucocyte function-associated antigen-1-dependent cellular conjugation by divalent cations. *Immunology* **81**, 120-126.
104. Dransfield, I., Cabanas, C., Craig, A. & Hogg, N. (1992). Divalent cation regulation of the function of the leucocyte integrin LFA-1. *J. Cell Biol.* **116**, 219-226.
105. Weitzman, J.B., Pasqualini, R., Takada, Y. & Menler, M.E. (1993). The function and distinctive regulation of the integrin VLA-3 in cell-adhesion, spreading, and homotypic cell-aggregation. *J. Biol. Chem.* **268**, 8651-8657.
106. van Kooyk, Y., Weder, P., Heije, K. & Figdor, C. (1994). Extracellular Ca²⁺ modulates leukocyte function-associated antigen-1 cell surface distribution on T-lymphocytes and consequently affects cell adhesion. *J. Cell Biol.* **124**, 1061-1070.
107. Du, X., Plow, E.F., Frelinger, A.L., III, O'Toole, T.E., Loftus, J.C. & Ginsberg, M.H. (1991). Ligands "activate" integrin $\alpha_{IIb}\beta_3$ (Platelet GPIIb-IIIa). *Cell* **65**, 409-416.
108. Lam, S.C.-T., Plow, E.F., Smith, M.A., Andrieux, A., Ryckwaert, J.-J., Marguerie, G. & Ginsberg, M.H. (1987). Evidence that arginyl-glycyl-aspartate peptides and fibrinogen gamma chain peptides share a common binding site on platelets. *J. Biol. Chem.* **262**, 947-950.

109. D'Souza, S.E., Ginsberg, M.H., Lam, S.C.-T. & Plow, E.F. (1988). Chemical cross-linking of arginyl-glycyl-aspartic acid peptides to an adhesion receptor on platelets. *J. Biol. Chem.* **263**, 3943-3951.
110. Kirchhofer, D., Languino, L.R., Rouslahti, E. & Pierschbacher, M.D. (1990). $\alpha 2\beta 1$ integrins from different cell types show different binding specificities. *J. Biol. Chem.* **265**, 615-618.
111. Conforti, G., Zanetti, A., Pasquali-Ronchetti, I., Quaglino, D.J., Neyroz, P. & Dejana, E. (1990). Modulation of vitronectin receptor binding by membrane lipid composition. *J. Biol. Chem.* **265**, 4011-4019.
112. Akiyama, S.K., Yamada, S.S. & Yamada, K.M. (1989). Analysis of the role of glycosylation of the human fibronectin receptor. *J. Biol. Chem.* **264**, 18,011-18,018.
113. Turner, M.L. (1992). Cell adhesion molecules: A unifying approach to topographic biology. *Biol. Rev.* **67**, 359-377.
114. de Fougères, A.R. & Springer, T.A. (1992). Intercellular adhesion molecule-3, a third adhesion counter-receptor for lymphocyte function associated antigen-1 on resting lymphocytes. *J. Exp. Med.* **175**, 185-190.
115. Williams, A.F. & Barclay, A.N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. *Annu. Rev. Immunol.* **6**, 381-405.
116. Hunkapiller, T. & Hood, L. (1989). Diversity of the immunoglobulin gene superfamily. *Adv. Immunol.* **44**, 1-63.
117. Yoshihara, Y., Oka, S., Ikeda, J. & Mori, K. (1991). Immunoglobulin superfamily molecules in the nervous system. *Neuroscience Research* **10**, 83-105.
118. Schiffer, M. (1973). Structure of a λ -type Bence-Jones protein at 3.5Å resolution. *Biochemistry* **12**, 4620-4631.

119. Poljak, R.J. & al., e. (1973). Three dimensional structure of the Fab fragment of a human immunoglobulin at 2.8Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3305-3310.
120. Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. in Protein Chem.* **34**, 167-339.
121. Williams, A.F. (1987). A year in the life of the immunoglobulin superfamily. *Immunol. Today* **8**, 298-303.
122. Kornblihtt, A.R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F.E. (1985). Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene. *EMBO J.* **4**, 1755-1759.
123. Bork, P. & Doolittle, R.F. (1992). Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA* **89**, 8990-8994.
124. Ruoslahti, E. (1988). Fibronectin and its receptors. *Ann. Rev. Biochem.* **57**, 375-413.
125. Pierschbacher, M.D. & Ruoslahti, E. (1984). Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature* **309**, 30-33.
126. Aota, S., Nagai, T., Olden, K., Akiyama, S.K. & Yamada, K.M. (1991). Fibronectin and integrins in cell adhesion and migration. *Bioch. Soc. T.* **19**, 830-835.
127. Erickson, H.P. & Bourdon, M.A. (1989). Tenascin - an extracellular matrix protein prominent in specialized embryonic tissues and tumors. *Annu. Rev. Cell Biol.* **5**, 71-92.
128. Furley, A.J., Morton, S.B., Manalo, D., Karagogeos, D., Dodd, J. & Jessell, T.M. (1990). The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth-promoting activity. *Cell* **61**, 157-170.

129. Burgoon, M.P., Grumet, M., Mauro, V., Edelman, G.M. & Cunningham, B.A. (1991). Structure of the chicken neuron-glia cell adhesion molecule, Ng-CAM: origin of the polypeptides and relation to the Ig superfamily. *J. Cell Biol.* **112**, 1017-1029.
130. Volkmer, H., Hassel, B., Wolff, J.M., Frank, R. & Rathjen, F.G. (1992). Structure of the axonal surface recognition molecule neurofascin and its relationship to a neural subgroup of the immunoglobulin superfamily. *J. Cell Biol.* **118**, 149-161.
131. Leahy, D.J., Hendrickson, W.A., Aukhil, I. & Erickson, H.P. (1992). Structure of a Fibronectin Type III Domain from Tenascin Phased by MAD Analysis of the Selenomethionyl Protein. *Science* **258**, 987-991.
132. Baron, M., Main, A.L., Driscoll, P.C., Mardon, H.L., Boyd, J. & Campbell, I.A. (1992). ¹H NMR assignment and secondary structure of the cell adhesion type III module of fibronectin. *Biochemistry* **31**, 2068-2073.
133. Main, A.L., Harvey, T.S., Baron, M., J., B. & Campbell, I.A. (1992). The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions. *Cell* **71**, 671-678.
134. Dickinson, C.D., Veerapandian, B., Dai, X.-P., Hamlin, R.C., Xuong, N., Ruoslahti, E. & Ely, K.R. (1994). Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J. Mol. Biol.* **236**, 1079-1092.
135. Cunningham, B.A., Hemperly, J.J., Murray, B.A., Prediger, E.A., Brackenbury, R. & Edelman, G.M. (1987). Neural cell adhesion molecules: Structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science* **236**, 799-806.
136. Barthels, D., Vopper, G., Boned, A., Cremer, H. & Wille, W. (1992). High degree of NCAM diversity generated by alternative RNA splicing in brain and muscle. *Eur. J. Neurosci.* **4**, 327-337.
137. Cole, G.J. & Akeson, R.A. (1989). Identification of a heparin binding domain of the neural cell adhesion molecule N-CAM using synthetic peptides. *Neuron* **2**,

1157-1165.

138. Rutishauser, U., Acheson, A., Hall, A.K., Mann, D.M. & Sunshine, J. (1988). The neural cell adhesion molecule (NCAM) as a regulator of cell-cell interactions. *Science* **240**, 53-57.
139. Hoffman, S. & Edelman, G.M. (1983). Kinetics of homophilic binding by E and A forms of the neural cell adhesion molecule. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5762-5766.
140. Letourneau, P.C. & Shattuck, T.A. (1989). Distribution and possible interactions of actin-associated proteins and cell adhesion molecules of nerve growth cones. *Development* **105**, 505-519.
141. Pollerberg, G.E., Burrige, K., Krebs, K.E., Goodman, S.R. & Schachner, M. (1987). The 180-kD component of the neural cell adhesion molecule N-CAM is involved in cell-cell contacts and cytoskeleton-membrane interactions. *Cell Tissue Res.* **250**, 227-236.
142. Doherty, P., Ashton, S.V., Moore, S.E. & Walsh, F.S. (1991). Morphoregulatory activities of NCAM and N-cadherin can be accounted for by G protein-dependent activation of L- and N-type neuronal Ca²⁺ channels. *Cell* **67**, 21-33.
143. Frei, T., von Bohlen und Halbach, F., Wille, W. & Schachner, M. (1992). Different extracellular domains of the neural cell adhesion molecule (N-CAM) are involved in different functions. *J. Cell Biol.* **118**, 177-194.
144. Snow, P.M., Zinn, K., Harrelson, A.L., McAllister, L., Schilling, J., Bastiani, M.J., Makk, G. & Goodman, C.S. (1988). Characterization and cloning of fasciclin I and fasciclin II glycoproteins in the grasshopper. *Proc. Natl. Acad. Sci. USA* **85**, 5291-5295.
145. Grenningloh, G., Bieber, A.J., Rehm, E.J., Snow, P.M., Traquina, Z.R., Hortsch, M., Patel, N.H. & Goodman, C.S. (1990). Molecular genetics of neuronal recognition in *Drosophila*: evolution and function of the immunoglobulin superfamily

- cell adhesion molecules. *Cold Spring Harbor Symposia on Quantitative Biology* **55**, 327-340.
146. Brümmendorf, T., Wolff, J.M., Frank, R. & Rathjen, F.G. (1989). Neural cell recognition molecule F11: homology with fibronectin type III and immunoglobulin type C domains. *Neuron* **2**, 1351-1361.
 147. Zuellig, R.A., Rader, C., Schroeder, A., Kalousek, M.B., von Bohlen, F., Osterwalder, T., Inan, C., Stoeckli, E.T., Affolter, H.U., Fritz, A., Hafen, E. & Sonderegger, P. (1992). The axonally secreted cell adhesion molecule, axonin-1: primary structure, immunoglobulin-like and fibronectin-type III-like domains and glycosyl-phosphatidylinositol anchorage. *Eur. J. Biochem.* **204**, 453-463.
 148. Gennarini, G., Cibelli, G., Rougon, G., Mattei, M.G. & Goidis, C. (1989). The mouse neuronal cell surface protein F3: a phosphatidylinositol-anchored member of the immunoglobulin superfamily related to chick contactin. *J. Cell Biol.* **109**, 775-788.
 149. Brümmendorf, T., Hubert, M., Treubert, U., Leuschner, R., Tárnok, A. & Rathjen, F.G. (1993). The axonal recognition molecule F11 is a multifunctional protein: specific domains mediate interactions with Ng-CAM and restrictin. *Neuron* **10**, 1-20.
 150. Kuhn, T.B., Stoeckli, E.T., Condrau, M.A., Rathjen, F.G. & Sonderegger, P. (1991). Neurite outgrowth on immobilized axonin-1 is mediated by a heterophilic interaction with L1(G4). *J. Cell. Biol.* **115**, 1113-1126.
 151. Zisch, A.H., D'Alessandri, L., Ranscht, B., Falchetto, R., Winterhalter, K.H. & Vaughan, L. (1992). Neuronal cell adhesion molecule contactin/F11 binds to tenascin via its immunoglobulin-like domains. *J. Cell Biol.* **119**, 203-213.
 152. Moos, M., Tacke, R., Scherer, H., Teplow, D., Früh, K. & Schachner, M. (1988). Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. *Nature* **334**, 701-703.

153. Prince, J.T., Alberti, L., Healy, P.A., Nauman, S.J. & Stallcup, W.B. (1991). Molecular cloning of NILE glycoprotein and evidence for its continued expression in mature rat CNS. *J. Neurosci. Res.* **30**, 567-581.
154. Grumet, M., Mauro, V., Burgoon, M.P., Edelman, G.M. & Cunningham, B.A. (1991). Structure of a new nervous system glycoprotein, NrCAM, and its relationship to subgroups of neural cell adhesion molecules. *J. Cell Biol.* **113**, 1399-1412.
155. Sonderegger, P. & Rathjen, F.G. (1992). Regulation of axonal growth in the vertebrate nervous system by interactions between glycoproteins belonging to two subgroups of the immunoglobulin superfamily. *J. Cell Biol.* **119**, 1387-1394.
156. Grumet, M. (1992). Structure, expression, and function of Ng-CAM, a member of the immunoglobulin superfamily involved in neuron-neuron and neuron-glia adhesion. *J. Neurosci. Res.* **31**, 1-13.
157. Schuch, U., Lohse, M.J. & Schachner, M. (1989). Neural cell adhesion molecules influence second messenger systems. *Neuron* **3**, 13-20.
158. Williams, E.J., Doherty, P., Turner, G., Reid, R.A., Hemperly, J.J. & Walsh, F.S. (1992). Calcium influx into neurons can solely account for cell contact dependent neurite outgrowth stimulated by transfected L1. *J. Cell Biol.* **119**, 883-892.
159. Lemmon, V., Farr, K.L. & Lagenaur, C. (1989). L1-mediated axon outgrowth occurs via a homophilic binding mechanism. *Neuron* **2**, 1597-1603.
160. Mauro, V.P., Krushel, L.A., Cunningham, B.A. & Edelman, G.M. (1992). Homophilic and heterophilic binding activities of Nr-CAM, a nervous system cell adhesion molecule. *J. Cell Biol.* **119**, 191-202.
161. Kadmon, G., Kowitz, A., Altevogt, P. & Schachner, M. (1990). Functional cooperation between the neural cell adhesion molecules L1 and N-CAM is carbohydrate dependent. *J. Cell Biol.* **110**, 209-218.

162. Kadmon, G., Kowitz, A., Altevogt, P. & Schachner, M. (1990). The neural cell adhesion molecule N-CAM enhances L1-dependent cell-cell interactions. *J. Cell Biol.* **110**, 193-208.
163. Bieber, A.J., Snow, P.M., Hortsch, M., Patel, N.H., Jacobs, J.R., Traquina, Z.R., Schilling, J. & Goodman, C.S. (1989). *Drosophila* Neuroglian: A member of the immunoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. *Cell* **59**, 447-460.
164. Simister, N.E. & Mostov, K.E. (1989). Cloning and expression of the neonatal rat intestinal Fc receptor, a major histocompatibility complex class I antigen homolog. *Cold Spring Harbor Symposia on Quantitative Biol* **LIV**, 571-580.
165. Simister, N.E. & Mostov, K.E. (1989). An Fc receptor structurally related to MHC class I antigens. *Nature* **337**, 184-187.

Chapter 2

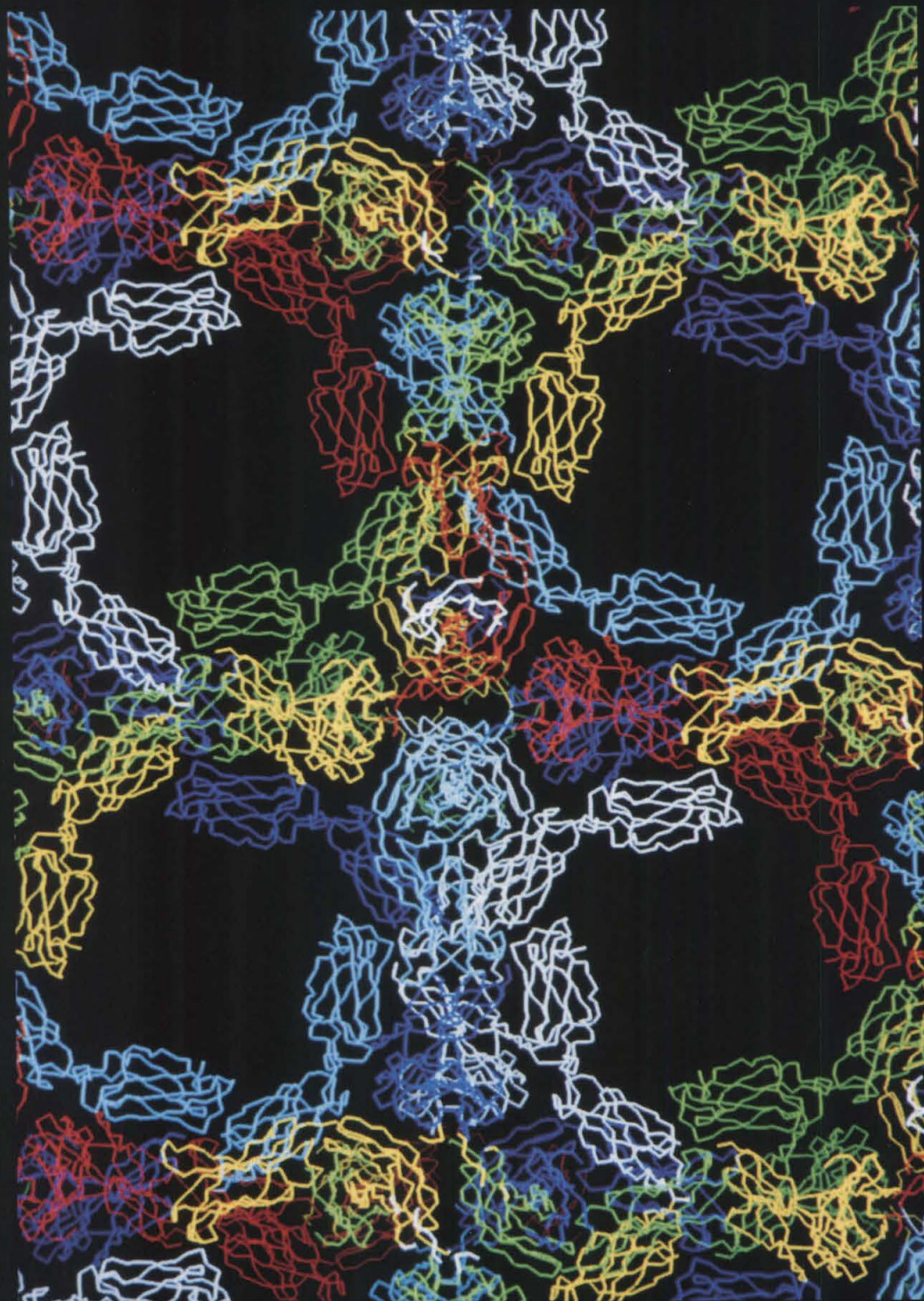
Crystal Structure of Tandem Type III Fibronectin Domains from *Drosophila* Neuroglian at 2.0Å

Figure 2-0 (following page) Neuron cover. A proteolytic fragment containing the first two fibronectin type III repeats from *Drosophila* neuroglian as it is arranged in the crystals (space group F432, $a = b = c = 241.8\text{\AA}$) used to solve its structure.

Neuron⁴⁹

Volume 12 Number 4

April 1994



Crystal Structure of Tandem Type III Fibronectin Domains from *Drosophila* Neuroglian at 2.0 Å⁵⁰

Andrew H. Huber,* Yu-mei Eureka Wang,[‡]
Allan J. Bieber,[‡] and Pamela J. Bjorkman**

*Division of Biology

†Howard Hughes Medical Institute
California Institute of Technology
Pasadena, California 91125

‡Department of Biological Sciences
Purdue University
West Lafayette, Indiana 47907

Summary

We report the crystal structure of two adjacent fibronectin type III repeats from the *Drosophila* neural cell adhesion molecule neuroglian. Each domain consists of two antiparallel β sheets and is folded topologically identically to single fibronectin type III domains from the extracellular matrix proteins tenascin and fibronectin. β bulges and left-handed polyproline II helices disrupt the regular β sheet structure of both neuroglian domains. The hydrophobic interdomain interface includes a metal-binding site, presumably involved in stabilizing the relative orientation between domains and predicted by sequence comparison to be present in the vertebrate homolog molecule L1. The neuroglian domains are related by a near perfect 2-fold screw axis along the longest molecular dimension. Using this relationship, a model for arrays of tandem fibronectin type III repeats in neuroglian and other molecules is proposed.

Introduction

Many cell-cell interactions in the nervous system are mediated by immunoglobulin gene superfamily members (Yoshihara et al., 1991). In neural cell adhesion molecules such as L1, TAG-1, contactin, fasciclin II, and neuroglian, tandem immunoglobulin-like domains are connected in series with two or more fibronectin type III (Fn-III) repeats. A number of such molecules have been identified in both vertebrates and invertebrates (Grenningloh et al., 1990). For example, the extracellular portion of neuroglian, a *Drosophila* neural recognition molecule, consists of six immunoglobulin repeats of the C2 type (Williams and Barclay, 1988) followed by five Fn-III repeats, a domain organization shared by the mammalian neural adhesion molecule L1 (Bieber et al., 1989). Neuroglian and L1 are 28% identical at the amino acid level and are suggested to share a common evolutionary ancestor (Bieber et al., 1989). Neuroglian is believed to function similarly to L1, which has been implicated in neural adhesion, neurite outgrowth, and axonal fasciculation (Hortsch and Goodman, 1991). Neuroglian acts as a homophilic adhesion molecule *in vitro*, and cells expressing the protein sort away from cell clusters expressing other *Drosophila* cell adhesion molecules

(A. Bieber et al., unpublished data). Mutant *Drosophila* embryos completely lacking neuroglian expression show relatively normal development of the central and peripheral nervous systems. However, this mutation is lethal, and the orientation and extent of contact among sensory neurons normally expressing the protein are abnormal, consistent with the proposed role of neuroglian in cell adhesion events (Grenningloh et al., 1990).

Although commonly present in neural adhesion molecules such as neuroglian, Fn-III domains were originally identified as a repeating motif of ~90 amino acids in the extracellular matrix (ECM) protein fibronectin. Some Fn-III domains in ECM proteins interact with integrins via an Arg-Gly-Asp (RGD) sequence motif (Hynes, 1990). An example of an interaction of Fn-III domains with a member of the immunoglobulin superfamily is provided by the observation that some of the Fn-III motifs of the ECM protein tenascin bind to contactin (Zisch et al., 1992). The interactions between the Fn-III domains of neural adhesion molecules and their ligands are less well characterized. However, the two Fn-III domains of the neural adhesion molecule N-CAM, whose extracellular region consists of five tandem immunoglobulin C2 repeats followed by two Fn-III domains, have been found to be sufficient for some of the functions of N-CAM, including spreading of neuronal cell bodies (Frei et al., 1992). Other studies mapping the functions of L1 to specific domains show that the first two Fn-III repeats of the molecule, in isolation, are capable of promoting neurite outgrowth, whereas the final three Fn-III repeats cause adherence of small cerebellar neurons (Appel et al., 1993). In addition to comprising parts of adhesion molecules and proteins of the ECM, Fn-III modules have been found in cytoplasmic muscle proteins, the extracellular regions of receptor protein kinases, receptor protein phosphatases, cytokine receptors, and prokaryotic enzymes that cleave carbohydrates (Bork and Doolittle, 1992). After a systematic screen of the protein sequence data base, it was estimated that the Fn-III motif occurs in about 2% of all animal proteins (Bork and Doolittle, 1992), indicating that it is a common structural motif.

The structures of single Fn-III domains from the ECM proteins tenascin (third Fn-III domain of tenascin [TnFn3]) and fibronectin (tenth Fn-III domain of fibronectin [FnFn10]) are known from X-ray crystallographic and nuclear magnetic resonance studies (Leahy et al., 1992; Baron et al., 1992; Main et al., 1992). These domains are related by statistically significant sequence identity and share a common fold, consisting of seven β strands forming two antiparallel β sheets, with conserved hydrophobic residues packing in their central cores. The topology is similar to that of an immunoglobulin constant region, except for the "sheet switching" of one β strand (Leahy et al.,

1992). A structural similarity between cytokine receptor domains and the Fn-III module was predicted by analysis of patterns of hydrophobic and hydrophilic residues (Bazan, 1990; Patthy, 1990), and the prediction was verified by the structure of the two extracellular domains of the human growth hormone receptor (De Vos et al., 1992). The Fn-III fold is also identical to the topology of β strands in the bacterial chaperonin PapD (Holmgren and Brändén, 1989) and the second domain of CD4 (Ryu et al., 1990; Wang et al., 1990). In these cases, however, the topological similarity to Fn-III modules is thought to have arisen by convergent evolution (Main et al., 1992).

We report here the 2.0 Å crystal structure of two tandem Fn-III domains from the *Drosophila* cell adhesion molecule neuroglian. Although identifiable as a repeat ~90–100 residues in length containing a conserved pattern of hydrophobic residues, the Fn-III modules of the neural adhesion proteins do not show statistically significant sequence similarity (defined as >15% or preferably >25% sequence identity between pairs of potentially related sequences; Doolittle, 1987) to ECM proteins. Nevertheless, the folding topology of each neuroglian domain is identical to that shared by the Fn-III domains of tenascin and fibronectin (Leahy et al., 1992; Baron et al., 1992; Main et al., 1992). The relative orientation of Fn-III domains in series, as occurs in the neural adhesion molecules and ECM proteins, was previously unknown. In the ECM protein fibronectin, Fn-III modules N-terminal to the RGD-containing Fn-III domain are required for maximal adhesive function (Aota et al., 1991; Nagai et al., 1991); thus, the interaction of Fn-III repeats in series is of considerable interest. We find that the neuroglian domains are related by a near dyad axis of symmetry with a translation along the pseudosymmetry axis. A sodium ion with approximate square pyramidal coordination is found at the interface between the two neuroglian domains, presumably stabilizing their relative orientation. An analysis of the arrangement of the two tandem neuroglian Fn-III modules allows a general model to be proposed for the arrangement of multiple Fn-III repeats in adhesion molecules, receptors, and ECM proteins. Together with structural information concerning immunoglobulin-like domains from the crystal structures of CD4 (Ryu et al., 1990; Wang et al., 1990; Brady et al., 1993) and CD2 (Jones et al., 1992), the model can be used to estimate the overall dimensions of cell adhesion molecules, which are predicted to be narrow (20–30 Å) molecules that can extend up to 370 Å from the cell.

Results

Expression of Neuroglian Fragments and Proteolysis of the Fn-III Repeats

Plasmids for the expression of soluble fragments of neuroglian were prepared as described in the Experimental Procedures. These secretion vectors were transfected into *Drosophila* S2 cells. Clonal cell lines

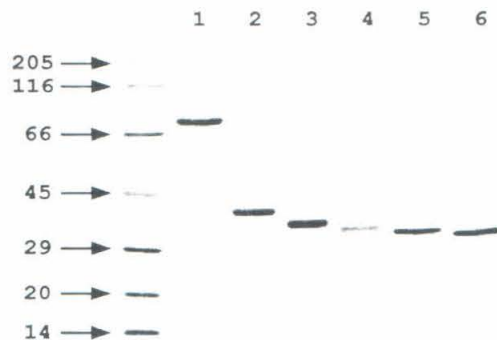


Figure 1. SDS-Polyacrylamide Gel Electrophoresis (15%) of Glycosylated and Deglycosylated Forms of the Chymotryptic Fragment of the Five Fn-III Repeat Protein

Lane 1, Purified five Fn-III repeat protein prior to proteolysis; lane 2, purified chymotrypsin NgFn1,2 fragment; lane 3, NgFn1,2 protein derived from washed crystals; lane 4, NgFn1,2 protein treated with TFMS; lane 5, NgFn1,2 protein treated with PNGase F; lane 6, crystallized NgFn1,2 protein treated with PNGase F. The slight difference in position of the protein species in lanes 5 and 6 may be due to residual amounts of harvest buffer in the washed crystal sample altering its gel mobility, or the possibility that the protein which crystallized represents a minor population of protein that is slightly smaller in mass than the majority of the starting protein. Arrows, molecular weight in kilodaltons.

secreting high levels of the fragments were identified and used to produce large quantities of the proteins for structural analyses.

Conditions were found that yielded crystals of the entire 160 kD extracellular portion of neuroglian, but these crystals did not diffract to sufficient resolution to allow a structure determination. Proteins composed of multiple repeats arranged in series often do not form well-ordered crystals because of the flexibility between domains (Kwong et al., 1990). Therefore, a systematic approach was used to find a portion of neuroglian that could form crystals capable of high resolution diffraction. A segment consisting of the five Fn-III repeats was next crystallized, but again, these crystals did not diffract to atomic resolution. Because a stable proteolytic fragment of fibronectin consisting of three Fn-III modules had been reported (Nagai et al., 1991), we next surveyed several proteases to identify one that could cleave the five Fn-III repeat molecule. Proteolytic cleavage with trypsin, chymotrypsin, or elastase yielded a fragment with an apparent molecular mass of 37.5 kD on SDS-polyacrylamide gel (Figure 1). Sequence analysis showed that the chymotryptic fragment started at residue 610, and the tryptic fragment started at 607, near the predicted beginning of the first Fn-III repeat (Bieber et al., 1989). Upon treatment under denaturing conditions with peptide N-glycosidase F (PNGase F), which cleaves N-glycosidic linkages, the apparent molecular mass on an SDS-polyacrylamide gel was reduced to 31.5 kD (Figure 1). To rule out the possibility that some of the mass of the enzymatically deglycosylated fragment was due

Table 1. Data Collection Statistics

Data	Native	Ethylmercuric Phosphate
Number of Crystals	1	1
Resolution (Å)	1.8	2.4
Number of measurements	355,936	80,531
Unique reflections	45,766	19,507
% Complete (resolution Å)	99 (2.0)	86 (2.6)
R_{merge}	0.050	0.079

$R_{\text{merge}} = \Sigma (|I - \langle I \rangle|) / (\Sigma \langle I \rangle)$, in which I is the intensity of a reflection, and $\langle I \rangle$ is the mean intensity for the reflection.

to the presence of O-linked sugars, the fragment was deglycosylated with trifluoromethanesulfonic acid (TFMS), which removes all but the O-glycosidically linked galactosamine and N-glycosidically linked glucosamine of a complex carbohydrate moiety. The TFMS treatment did not reduce the apparent molecular mass beyond the mass reduction achieved by enzymatic deglycosylation, suggesting that the fragment contained only N-linked carbohydrates.

Crystallization and Structure Determination

Single well-ordered crystals of the chymotryptic fragment were grown and characterized as space group F432 ($a = b = c = 241.79$ Å) with one molecule per asymmetric unit. Surprisingly, SDS-polyacrylamide gel electrophoresis of the protein from washed crystals revealed an apparent molecular mass of 34 kd, which was smaller than the mass of the protein used in the crystallization (Figure 1). PNGase F treatment of protein from washed crystals reduced the apparent molecular mass to ~31 kd, similar to the deglycosylated mass of the protein used in the crystallizations, suggesting that a less heavily glycosylated subset of the protein had crystallized out of a mixture of heterogeneously glycosylated species.

Protein from washed crystals was analyzed by matrix-assisted laser desorption mass spectroscopy. The mass of the protein species that crystallized was determined to be 26,102 daltons (± 65 daltons), 8 kd lower than the apparent molecular mass determined by

SDS-polyacrylamide gel electrophoresis. These results are consistent with the crystallization of a proteolytic fragment consisting of the first two Fn-III repeats and ~3,000 daltons of carbohydrate, compatible with utilization of two of the three potential sites of N-linked glycosylation within these domains (Bieber et al., 1989). Native and derivative data sets were collected at the Cornell High Energy Synchrotron Source (CHESS; Table 1) from cryopreserved crystals (Hope, 1990). A 2.6 Å solvent-flattened (Wang, 1985) electron density map calculated using phases derived from a single isomorphous derivative (Table 2) was of high quality (Figure 2A), allowing the complete tracing of 205 residues of the polypeptide chain. The model has now been refined to an R value of 20.2% for data between 5 and 2.0 Å, with good geometry (Table 3).

The Neuroglian Fn-III Motifs Have Similar Structures

The structure of the proteolytic fragment of the neuroglian Fn-III repeats (hereafter referred to as NgFn1,2) consists of two Fn-III structural motifs arranged in series at an obtuse angle with respect to each other (Figure 3A). The overall fold of each Fn-III domain is identical to that reported for the structures of single Fn-III domains from human tenascin and fibronectin (Baron et al., 1992; Leahy et al., 1992; Main et al., 1992). This fold is composed of seven β strands arranged in two antiparallel β sheets of four and three β strands. The seven strands are labeled A, B, C, C', E, F, and G, and the two sheets are composed of strands A-B-E and C'-C-F-G (Figures 3B and 3C). The Fn-III fold is similar to the fold of an immunoglobulin constant domain, except for the sheet switching of strand C', which hydrogen bonds with strand C in the type III domain rather than with strand E as observed in the immunoglobulin domain (Leahy et al., 1992; Main et al., 1992).

In the first neuroglian Fn-III domain (NgFn1), a disulfide bond joins residue 625 of strand A to residue 706 of strand G. Although the disulfide bond connects the two β sheets of the Fn-III domain, it bears no other relationship to the type of disulfide bond connecting β sheets in immunoglobulin variable and constant domains and in domains of members of the immuno-

Table 2. Phasing Statistics

	Resolution (Å)								
	25.0-2.60	11.90	7.87	5.88	4.70	3.91	3.35	2.93	2.60
Mean figure of merit									
Acentric	0.31	0.52	0.59	0.56	0.49	0.39	0.32	0.25	0.18
Reflections (No.)	14,340	93	348	735	1,276	1,948	2,682	3,386	3,872
Centric	0.45	0.40	0.56	0.62	0.54	0.45	0.44	0.38	0.29
Reflections (No.)	2,304	73	169	246	314	363	398	381	360
Phasing power (acentric)	1.3	1.3	1.9	2.1	1.8	1.3	1.2	1.1	1.0
R_{cullis} (acentric)	0.75	0.7	0.63	0.54	0.67	0.80	0.79	0.82	0.88

Phasing power = $\langle F_H \rangle / E_{\text{iso}}$, the rms heavy atom structure factor amplitude divided by the residual isomorphous lack of closure error; $R_{\text{cullis}} = \Sigma |F_H(\text{obs}) - F_H(\text{calc})| / \Sigma F_H(\text{obs})$, in which $F_H(\text{obs})$ is the observed heavy atom structure factor amplitude for a reflection and $F_H(\text{calc})$ is the calculated amplitude.

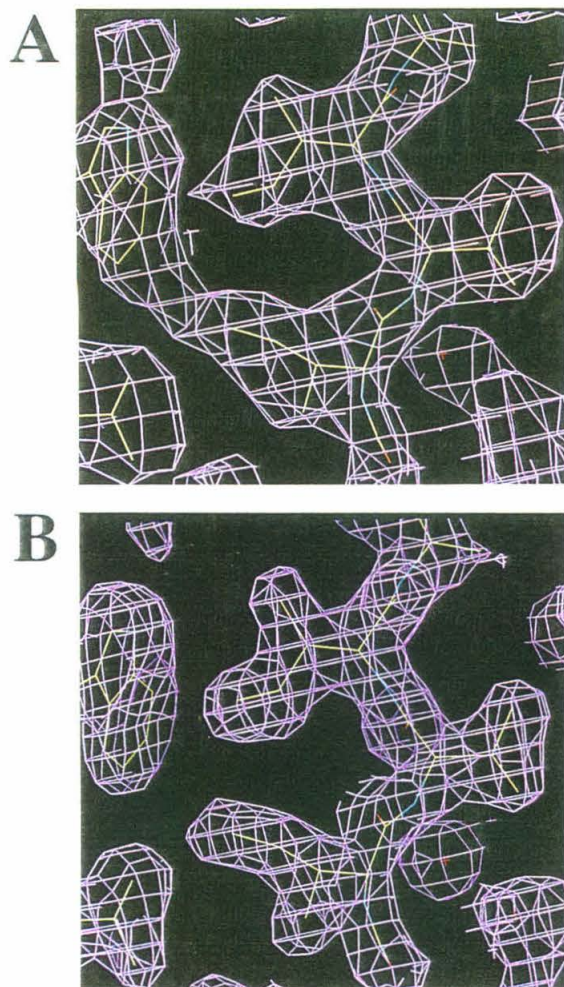


Figure 2. Initial and Refined Electron Density Maps
(A) Part of the 2.6 Å initial electron density map obtained by SIRAS and solvent flattening used to trace the polypeptide chain (see Experimental Procedures). The region shown is in the vicin-

globulin superfamily, which connects strands B and F. The second Fn-III repeat in neuroglian (NgFn2) does not contain a similar disulfide bond, and cysteine residues are not found in the comparable positions in sequences of the other neuroglian Fn-III domains (Bieber et al., 1989) or the Fn-III domains in tenascin (Leahy et al., 1992), fibronectin (Main et al., 1992), or human growth hormone receptor (De Vos et al., 1992).

Two potential N-linked glycosylation sites (Figure 3C) in NgFn1 are utilized. Clear density is seen for two N-acetylglucosamine residues attached to Asn-652 and a single N-acetylglucosamine attached to Asn-683. The two glycosylation sites are located on the adjacent β strands C and F of the four-stranded face of NgFn1 (Figure 3B).

Each of the neuroglian Fn-III domains is ~100 amino acids long, as compared with 91 residues in the TnFn3 and FnFn10 structures (Leahy et al., 1992; Main et al., 1992). The additional neuroglian residues are located at the C-terminus of each domain and in the loops connecting strands B and C. Although some individual β strands have different conformations in the neuroglian and tenascin Fn-III modules, the hydrophobic residues conserved in most Fn-III domains occupy analogous positions in the TnFn3 and NgFn1,2 domain interiors. NgFn1 and NgFn2 are superimposed upon each other and upon the TnFn3 structure in Figures 4A–4C. Overall, 61 carbon- α atoms of NgFn1 superimpose within 2.5 Å of their counterparts in TnFn3 with a root-mean-square (rms) error of 1.3 Å, with the comparable numbers being 67 carbon- α atoms and an rms error of 1.2 Å for the superposition of NgFn2 and

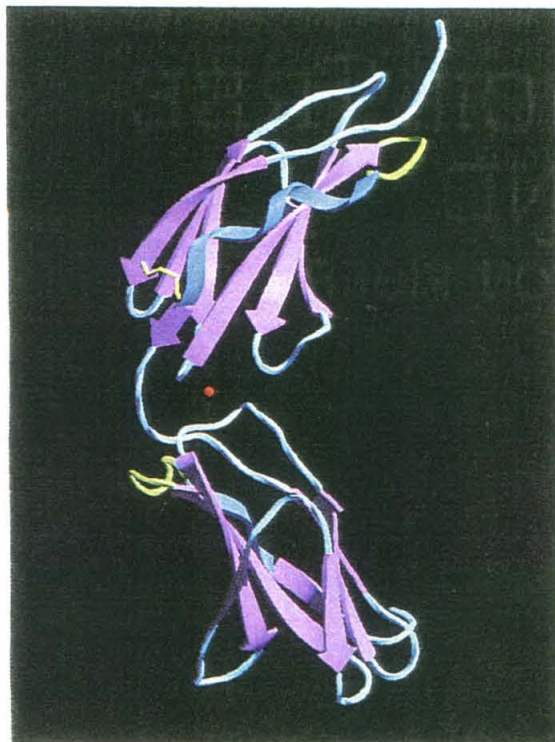
ity of three residues from strand E of NgFn2 (Ile-775, Val-776, Ile-777; top to bottom), contoured at the 1.0 σ level. Carbonyl oxygens and some water molecules (marked with a red cross) were identifiable in this initial map. The model shown is that obtained after the final refinement cycle.
(B) Corresponding section of a 2.1 Å electron density ($2F_{\text{obs}} - F_{\text{calc}}$) map contoured at the 1.0 σ level.

Table 3. Refinement Statistics

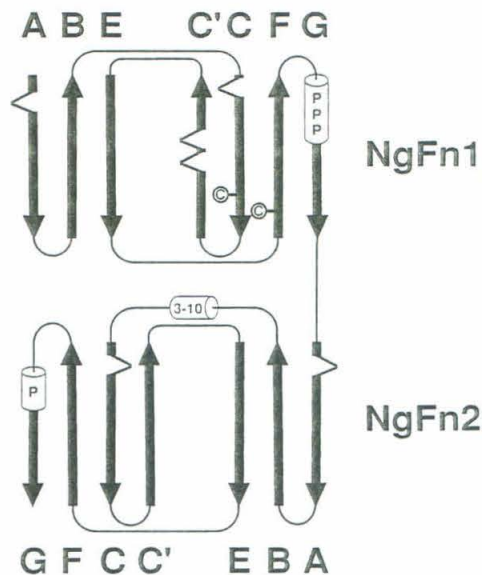
Resolution (Å)	Fraction Complete	R_{cryst}		R_{free}
		Fo > 0* σ_F	Fo > 3* σ_F	Fo > 0* σ_F
10.0–5.00	0.99	0.343	0.343	0.336
5.00–3.62	1.0	0.187	0.186	0.221
3.62–3.09	1.0	0.189	0.187	0.220
3.09–2.77	1.0	0.205	0.200	0.232
2.77–2.55	1.0	0.216	0.209	0.253
2.55–2.39	1.0	0.217	0.206	0.237
2.39–2.27	0.99	0.208	0.199	0.242
2.27–2.16	0.99	0.209	0.196	0.255
2.16–2.07	0.99	0.222	0.205	0.265
2.07–2.00	0.99	0.237	0.202	0.275
5.00–2.00		0.202	0.195	0.235
10.0–2.00		0.232	0.227	0.257

$R_{\text{cryst}} = \sum_a (|F_o - F_c|) / (\sum_a F_o)$, in which F_o and F_c are observed and calculated structure factor amplitudes and a is (refinement-set reflections);
 $R_{\text{free}} = \sum_t (|F_o - F_c|) / (\sum_t F_o)$, in which t is (free-set reflections); σ_F is the estimated SD for each structure factor amplitude.

A



B



C

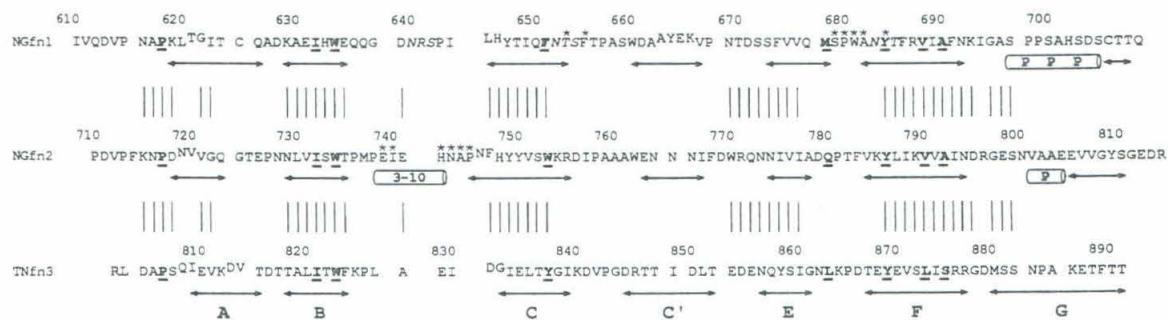


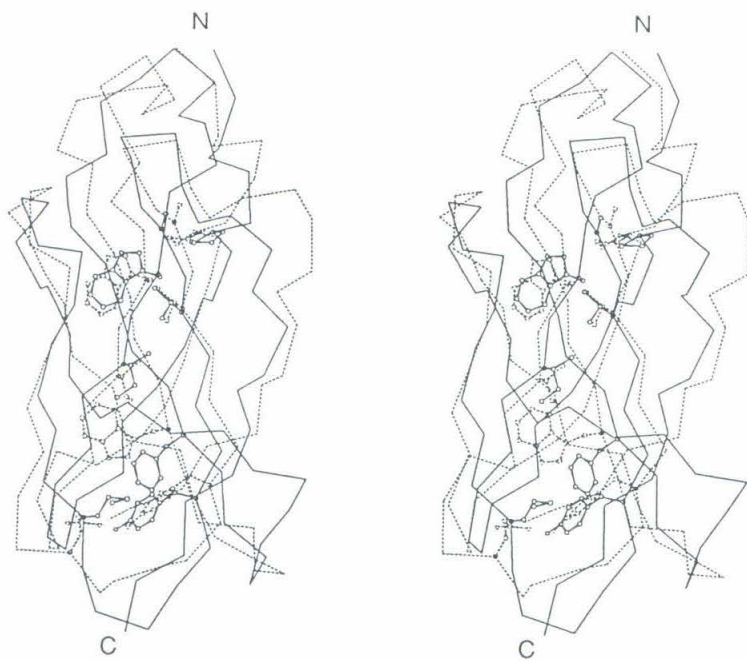
Figure 3. The Structure of NgFn1,2 and Sequence Alignment with TnFn3

(A) Ribbon diagram of the two Fn-III domains of neuroglial, showing β strands as magenta arrows, polyproline helical segments in dark blue, a disulfide bond in yellow, and the metal ion at the domain interface as a red sphere. The F to G loops of each domain (location of the RGD sequence in ECM proteins) are highlighted in green. This diagram was generated using the program SETOR (Evans, 1993).

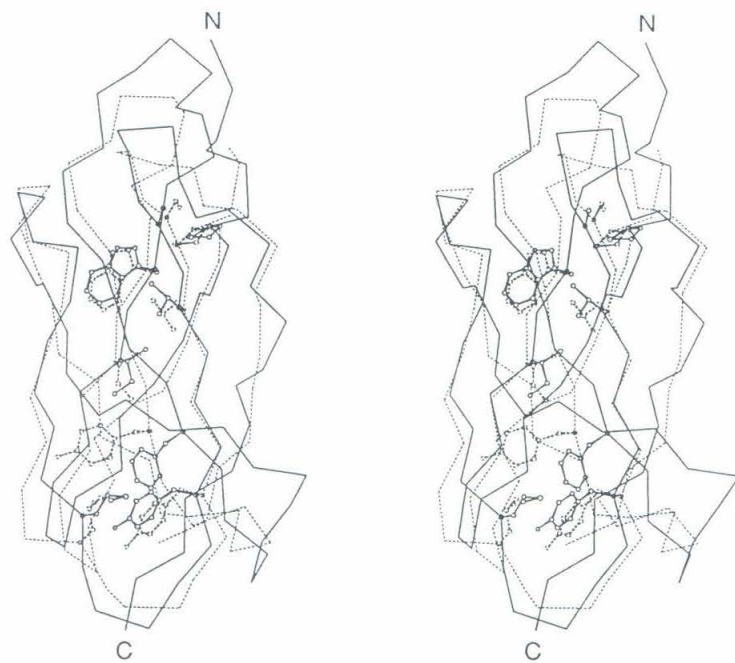
(B) Schematic diagram of the folding topology of the NgFn1,2 fragment. Within a single Fn-III domain, strands A, B, and E form one β sheet, whereas strands C', C, F, and G form the other. Approximate positions of β bulge elements are indicated by a break in the strand arrow, utilized N-linked glycosylation moieties by the letter C, and positions of left-handed polyproline II helix are indicated by cylinders with each helical turn denoted by the letter P. A single turn of 3_{10} helix is denoted by a cylinder marked 3-10.

(C) Sequence alignments of NgFn1, NgFn2, and TnFn3. These domains were aligned after superposition of their structures as described in the legend to Figure 4. The structural cores of residues whose carbon- α atoms are less than 2.5 Å away from corresponding residues in the other domains are marked by vertical lines. Residues commonly conserved in Fn-III domains are underlined and shown in bold, residues involved in β bulges are shown as superscripts, and N-linked glycosylation sites are in italics. Positions of the β strands in each domain are shown below its sequence, with the strand positions for TnFn3 taken from Leahy et al. (1992). Residues involved in interdomain contacts in the NgFn1,2 structure are marked with an asterisk (defined as a residue with any atom within 4 Å of an atom from a residue within the partner domain).

A



B



C

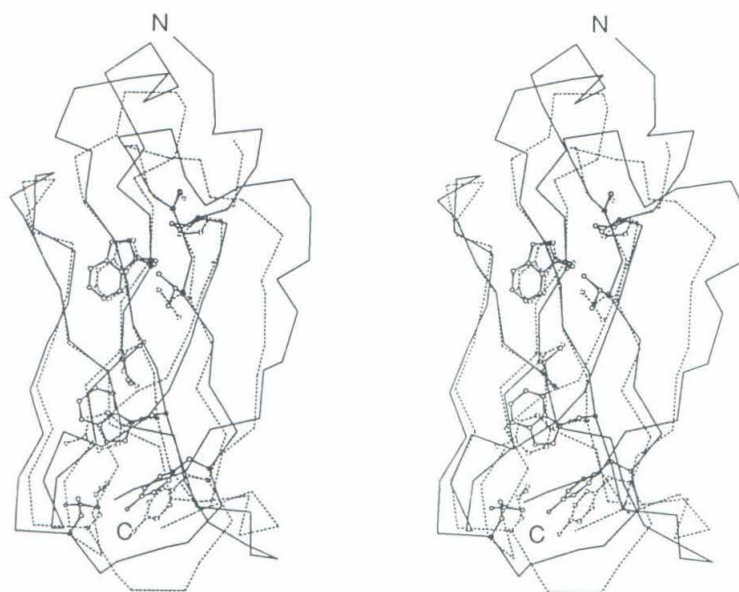


Figure 4. Superposition of NgFn1, NgFn2, and TnFn3

The carbon- α backbone of each domain following superposition is shown in stereo. Side chains of residues that are commonly conserved in Fn-III domains (Figure 3C, underlined and bold) are also shown. The domains were originally aligned manually using interactive graphics. Carbon- α atoms that superimposed well were used in a least squares alignment, and those that superimposed within 2.5 Å were used in a subsequent calculation to generate the alignments shown. N- and C-termini are indicated for the neuroglian domains in each stereoplot. These stereoplots were generated using Insight II from Biosym Technologies of San Diego.

(A) NgFn1 (solid) superimposed upon NgFn2 (dashed).

(B) NgFn1 (solid) superimposed upon TnFn3 (dashed).

(C) NgFn2 (solid) superimposed upon TnFn3 (dashed).

TnFn3. Each of the neuroglian domains is structurally more similar to TnFn3 than to its partner domain, as evidenced by the overall rms deviation of 1.6 Å for the superposition of 50 carbon- α pairs from the two neuroglian domains. The domains superimpose surprisingly well, considering the lack of significant sequence identity between the two neuroglian domains themselves and between either neuroglian domain and TnFn3 (16.5% identity between NgFn1 and NgFn2; 13% identity between TnFn3 and NgFn1; 9% identity between TnFn3 and NgFn2; see Figure 3C).

A comparison of the carbon- α atoms used in the alignment of the NgFn1, NgFn2, and TnFn3 domains yields a structural core of 42 residues that contains carbon- α atoms found to be within 2.5 Å of each other in all three domains (marked with vertical lines in Figure 3C). Most of these residues are contributed by strands B, E, C, and F, the four strands that interact to form the hydrophobic core of each domain and that contain most of the highly conserved hydrophobic residues commonly found in Fn-III domains. A portion of the residues in the F to G loop is also part of the structural core relating the two neuroglian domains to TnFn3. This loop is the location of the RGD adhesive sequence in TnFn3 and FnFn10 (Leahy et al., 1992; Main et al., 1992). The overall rms deviation between these core residues is 1.4 Å for the superposition of NgFn1 and NgFn2, 1.2 Å for NgFn1 and TnFn3, and 1.0 Å for NgFn2 and TnFn3.

Non- β Sheet Secondary Structure

The residues from strands A, C', and G were difficult to align in the NgFn1,2 and TnFn3 structures primarily

because of the interruption of regular β strand geometry by the insertion of other elements of secondary structure, such as β bulges (Richardson et al., 1978) and polyproline II helices (Adzhubei and Sternberg, 1993). Strand C' of NgFn1 contains adjacent β bulges, a classic bulge followed by a wide bulge, as defined by Richardson et al. (1978) (Figures 3B and 3C). The bulges make it difficult to superimpose NgFn1 strand C' on the C' strand of NgFn2 or TnFn3, which lacks similar bulges. However, the positions of two other β bulges are loosely conserved between the three domains even though the type of bulge is not (Figures 3B and 3C). β bulges are often found at active sites of proteins to orient side chains in needed directions and also serve as a mechanism for accommodating insertional mutations without completely disrupting β sheets (Richardson et al., 1978). A high number of β bulge secondary structural elements in the Fn-III modules of other proteins may contribute to their lack of statistical primary sequence similarity.

The NgFn1,2 structure also contains several examples of polyproline II helix, a structural motif in globular proteins that is characterized as a left-handed helix with a three residue repeat (Adzhubei and Sternberg, 1993). The main chain dihedral angles of polyproline II helices cluster in Φ, Ψ space around $-75^\circ, 145^\circ$, and these helices are found in polymers of proline, as part of a triple helix in the fibrous protein collagen and as short stretches of single helix in other proteins that usually, but not always, include the residue proline. This motif generally occurs on the protein surface and tends to be among the more mobile parts of the protein, because it has few main chain hydrogen bonds

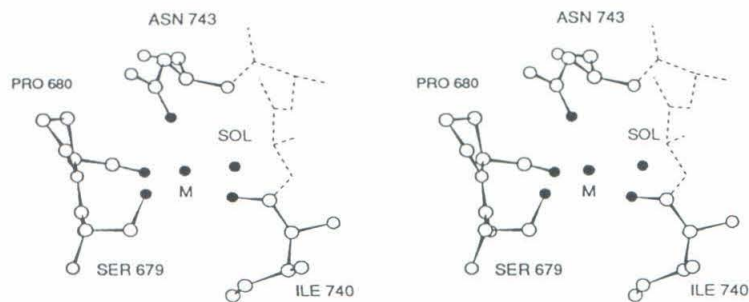


Figure 5. Environment around the Sodium Ion-Binding Site

The five oxygen ligands of the metal ion (M) are shown as solid spheres, and other atoms of the ligand residues are shown as open spheres. The residue number and identity of each ligand are indicated. SOL refers to the oxygen atom of an ordered water molecule. Dashed lines represent the intervening polypeptide chain between Ile-740 and Asn-743. This stereoplot was generated using Insight II from Biosym Technologies of San Diego.

with the rest of the protein (Adzhubei and Sternberg, 1993). Most of what was previously identified as strand G in TnFn3 (Figure 3C; Leahy et al., 1992) is three turns of polyproline II helix in NgFn1. This helix comprises at least nine amino acids (Ser-697–Ser-705), thus representing one of the longest single polyproline II helices yet found (Adzhubei and Sternberg, 1993). The polyproline II helix contains three serines which are evenly spaced (n , $n + 3$, and $n + 6$), such that the nitrogen and side chain hydroxyl of each serine hydrogen bond with the main chain of strand F. Each of the serine side chain hydroxyls also hydrogen bonds with polyproline II helix main chain carbonyl oxygens ($n + 1$, $n + 4$, and $n + 7$). The helix terminates and regular β strand hydrogen bonding resumes at Cys-706, which forms a disulfide bond with Cys-625. Examination of the Φ, Ψ angles of the comparable strands of NgFn2 and TnFn3 shows they are also interrupted by polyproline II structure shortly after the β hairpin turn connecting strands F and G.

The first serine in the NgFn1 polyproline II helix is conserved in NgFn2 and TnFn3 and is the first residue in the non- β strand structure after the F to G loop. In TnFn3, residues Ser-882–Pro-884 form a single non-ideal turn of polyproline II helix in which Ser-882 forms main chain–main chain and side chain–main chain hydrogen bonds analogous to those formed by Ser-697 in NgFn1. In NgFn2, non- β strand structure extends from Ser-799 to Glu-804, with residues Val-801 to Glu-804 forming one turn of polyproline II helix.

The F to G Loop

The adhesive properties of the ECM proteins that contain Fn-III modules are in part conferred by a portion of the module containing an RGD sequence (Pierschbacher and Ruoslahti, 1984). The RGD sequences of TnFn3 and FnFn10 are located in a β hairpin loop between strands F and G (Leahy et al., 1992; Main et al., 1992). The TnFn3 RGD loop is shorter by four amino acids than the FnFn10 loop and does not mediate interactions with the vitronectin receptor unless the TnFn3 domain is isolated from the rest of tenascin (Leahy et al., 1992). The function, if any, of the comparable loop in the Fn-III modules of neural adhesion molecules is unknown. None of the Fn-III modules in neuroglian or most other cell adhesion molecules (Bieber et al., 1989; Bork and Doolittle, 1992; P. Bork,

personal communication) contain RGD sequences. In all five Fn-III repeats in neuroglian, the strand F to G loop is shorter than the “active” loop of FnFn10, which mediates adhesive interactions in the whole fibronectin protein (Hynes, 1990). In fact, the F to G loop of each neuroglian Fn-III domain is a β hairpin turn and is the same size as the RGD loop in TnFn3 (Figure 3C). The F to G loop is the only loop that is spatially and structurally similar in the two domains of the NgFn1,2 structure. Although the β hairpin turn between strands F and G of NgFn2 is near the interdomain interface, it is still solvent accessible because NgFn1 is tilted away from this portion of the second domain. The F to G loop was reported to be mobile in the nuclear magnetic resonance structure of FnFn10 (Main et al., 1992); however, a crystallographic temperature factor analysis (data not shown) shows that these loops are well defined in both neuroglian domains.

Metal-Binding Site and Interdomain Contacts

Contacts between the two Fn-III modules in the NgFn1,2 structure are confined to the loops connecting β strands at the bottom of NgFn1 and the top of NgFn2. The E to F loop of the first domain and the B to C loop from the second domain form the majority of the interface, including a metal-binding site. The metal cation has approximate square pyramidal coordination geometry with a water molecule, side chain oxygens from Ser-679 (E to F loop, NgFn1) and Asn-743 (B to C loop, NgFn2), and carbonyl oxygens from Ile-740 (B to C loop, NgFn2) and Pro-680 (E to F loop, NgFn1) as ligands (Figure 5). There is no electron density corresponding to a sixth ligand in the position expected for a site with octahedral coordination geometry. The average ligand to metal distance after refinement is 2.23 Å, significantly smaller than the 2.7–2.8 Å distance published for potassium, but between the expected ligand to metal distances for magnesium (2.1 Å) and sodium or calcium (2.4 Å) (Yamashita et al., 1990; Glusker, 1991). Refinement of the cation occupancy and examination of difference Fourier maps calculated after soaking crystals in solutions of EDTA, calcium, or sodium (described in the Experimental Procedures) suggest that the metal is a sodium ion rather than a magnesium or calcium ion.

The remainder of the domain interface is predominantly hydrophobic and includes contributions from

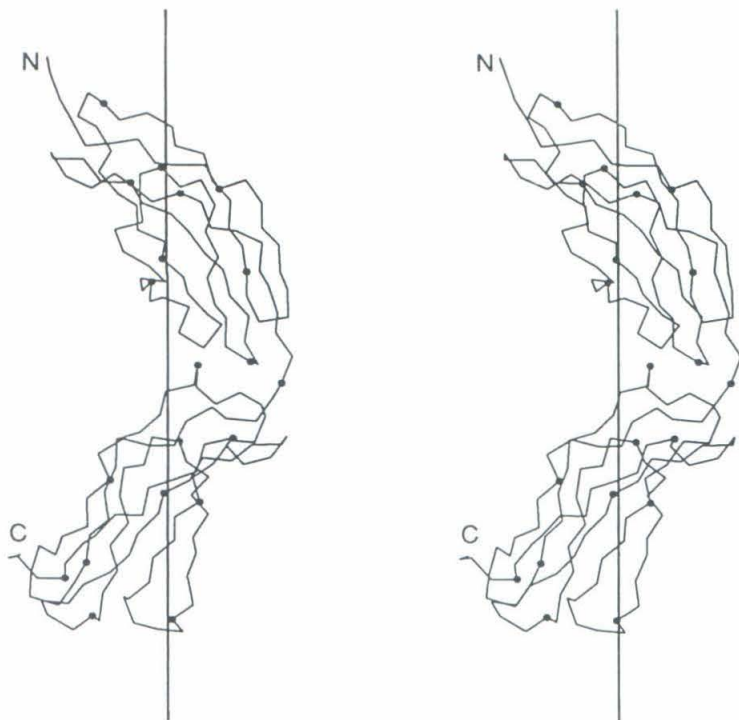


Figure 6. Location of the Pseudosymmetry Axis Relating NgFn1 and NgFn2

The carbon- α backbone of NgFn1,2 is shown in stereo with solid circles marking every tenth residue. The line indicates the position of the approximate 2-fold screw axis relating the two domains. The relationship between domains was calculated by superimposing NgFn1 upon NgFn2 as described in the legend to Figure 4 and is a 175° rotation followed by a 37.7 \AA translation along the rotation axis. This stereoplot was generated using Insight II from Biosym Technologies of San Diego.

the C to C' loop of NgFn1. Contact residues are marked with an asterisk in Figure 3C. Most of the contact residues in NgFn2 are in the B to C loop, which contains approximately five more residues than the analogous loops in the FnFn10 (Main et al., 1992) and TnFn3 domains (Figure 3C). This loop has a different structure in the two neuroglial domains, forming a turn of 3_{10} helix in NgFn2.

The interface between domains is stabilized by some 830 \AA^2 of total buried surface area, calculated by comparing the solvent-accessible surface area of NgFn1,2 with that of the separated domains. Similar calculations have been done for other structures consisting of tandem domains arranged in series: $620\text{--}880 \text{ \AA}^2$ are buried in the association of the first two domains of CD4 (Ryu et al., 1990; Jones et al., 1992), 950 \AA^2 are buried between the third and fourth CD4 domains (Brady et al., 1993), and 400 \AA^2 are buried between the two extracellular domains of CD2 (Jones et al., 1992).

Arrangement of the Two Fn-III Modules in the Neuroglial Fragment

The relative orientation of the two Fn-III domains in the crystal structure of NgFn1,2 is likely to be the same as what would be observed in solution for the following reasons. The resistance of NgFn1,2 to further proteolysis by three different proteases suggests that the fragment is a compact, stable structure without excessive flexibility between the domains. The observed domain interface is largely hydrophobic, implying

that it would be buried in the protein in solution. The surface area buried between adjacent domains is similar to that seen for other proteins with domains arranged in tandem. Finally, the presence of a metal-binding site at the domain interface strongly suggests that the orientation of the two domains is not merely a result of crystalline lattice contacts.

Each Fn-III domain in the neuroglial fragment can be roughly described as a cylinder of radius $\approx 10 \text{ \AA}$ and length $\approx 40 \text{ \AA}$. The two domains meet at an angle of $\sim 120^\circ$, as defined by the orientation of the two axes that would describe each cylinder. Because the two domains are tilted relative to each other, the length of the two domain fragment is $\sim 70 \text{ \AA}$, as opposed to $\sim 80 \text{ \AA}$, the predicted length of a linear two domain fragment. A tilted relationship of successive Fn-III repeats in the hexabrachion arm of human tenascin was proposed to reconcile the 36 \AA length of TnFn3 in the crystal structure to the 32 \AA repeat spacing observed in electron microscopic studies of human tenascin (Leahy et al., 1992). It was also proposed that alternative Fn-III domains would be rotated by $\sim 180^\circ$ to create an approximate 2-fold screw axis relating successive domains (Leahy et al., 1992). After a least squares superposition of the first NgFn1 upon NgFn2, the rotation and translation that relate the two domains are a 175° rotation followed by a 37.7 \AA translation along the rotation axis (Figure 6). This relationship is close to an exact 2-fold screw axis, in accordance with the prediction (Leahy et al., 1992). One consequence of the near dyad axis of rotation is that

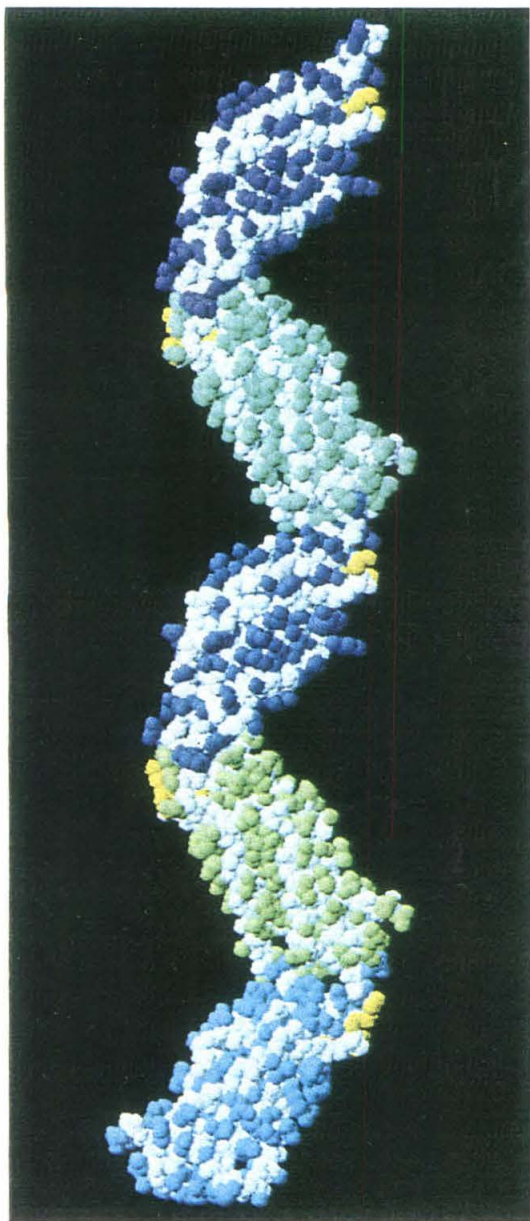


Figure 7. Proposed Model for the Alignment of Fn-III Domains in Series

Space-filling model in which main chain atoms in all domains are gray, and side chain atoms of each domain are colored differently than side chains of the neighboring domain. The F to G loop in each domain is highlighted in yellow. The model was generated by applying the rotation and translation that relates NgFn1 to NgFn2 three times in succession to generate the predicted positions of the final three Fn-III domains, assuming they are oriented with respect to each other with a similar relationship as NgFn1 and NgFn2. The result is a long fiber with a zig-zag repeat of two Fn-III units. Assuming other proteins containing multiple Fn-III repeats arranged in series are oriented in a similar manner, this model can be used as a first order approximation of the structures of the Fn-III domains in neural CAMs and in ECM proteins. This diagram was generated using the program SETOR (Evans, 1993).

the loop between strands F and G (highlighted in Figure 3A), which contains the RGD adhesive sequence in some Fn-III domains, would fall on opposite sides of the fiber on each successive repeat.

The relationship of the tandem domains in the NgFn1,2 structure is closer to being an exact dyad axis than the rotation relating pairs of domains in other relevant structures. In these other cases, the relationship between domains is a rotation of between 140° and 160° followed by a translation between 25 Å and 32 Å. For example, the first two domains of CD4 are related by $\sim 160^\circ$ (calculated using entry 1CD4 in the Protein Data Bank; Bernstein et al., 1977), as are the final two domains of CD4 (Brady et al., 1993). The two domains of PapD are related by 140° (calculated using entry 3DPA in the Protein Data Bank; Bernstein et al., 1977) and are oriented with a V shape creating a crevice thought to be the interaction area with pili subunits (Holmgren and Brändén, 1989). The extracellular domains of the human growth hormone receptor are also related by a 140° rotation (calculated using coordinates obtained from A. De Vos), and the relative domains are arranged almost perpendicular to each other, so that the dimeric molecule has a T shape (De Vos et al., 1992). Thus, molecules containing tandem repeated domains show a number of different relationships between successive domains. Of the above examples for which structural information is known, only CD4 contains more than two repeats.

The rotation axis relating NgFn1 and NgFn2 is approximately parallel to the longest dimension of the fragment (Figure 6), and successive application of the same rotation and translation along this axis yields a zig-zag arrangement of tandem Fn-III repeats that would ultimately form a straight rod. A computer-generated model of five Fn-III repeats (Figure 7) was created by three applications of the 175° rotation and 37.7 Å translation to the coordinates of the two domain structure and serves as a first order approximation of the structure of the five successive Fn-III domains in the extracellular portion of neuroglialin. If the relationship between domains depicted in Figure 7 is general for other proteins containing multiple Fn-III repeats, such as receptors or adhesive ECM proteins, the Fn-III repeat portions of such molecules would also be straight rods, with a zig-zag repeat unit of two Fn-III modules. Because the rotations that relate successive repeats in CD4, PapD, and human growth hormone receptor are not dyads, when the same procedure is used to create models of these proteins consisting of five repeated units, the resulting models (data not shown) do not have the regular appearance of the model shown in Figure 7.

Discussion

We have described the structure of the two N-terminal Fn-III repeats (NgFn1,2) from *Drosophila* neuroglialin, a neural CAM. NgFn1,2 is the only structure of Fn-III repeats from an adhesion molecule and also represents

the first structure in which the arrangement of tandem Fn-III repeats in an adhesion protein can be examined. The structures of single Fn-III repeats from the ECM proteins tenascin (TnFn3) and fibronectin (FnFn10) have been reported previously (Baron et al., 1992; Leahy et al., 1992; Main et al., 1992). Although the Fn-III motifs in neuroglian are related to the repeats in TnFn3 and FnFn10 by only a pattern of conserved hydrophobic residues and do not have detectable overall primary sequence similarity, both neuroglian domains have an identical topology to previously reported repeats. The tenascin, fibronectin, and neuroglian Fn-III domain structures are primarily composed of β sheet secondary structure, but the neuroglian domains contain an unusual number of β bulge elements that disrupt regular β sheet formation. The locations of some of these bulges are in common with those found in TnFn3. In addition, we find a segment of left-handed polyproline II helix in both neuroglian domains, as well as in TnFn3, so that the polypeptide chain following the F to G loop is more accurately described as a helix, rather than as a β strand.

The relative orientation of the two neuroglian Fn-III domains is stabilized by the binding of a sodium ion at the interface between the domains. Each domain contributes main chain and side chain ligands for the cation, which has approximate square pyramidal coordination geometry. An important role of metal ions in nonenzymatic proteins is to enhance the protein's structural stability in the conformation required for biological function (Glusker, 1991). Sodium- and potassium-binding sites have recently been observed in the structure of dialkylglycine decarboxylase (Toney et al., 1993). Both ions are too distant from the active site to be directly involved in catalysis, but instead play a role in maintaining the enzyme in the correctly folded conformation, analogous to the role we propose for the sodium ion in the NgFn1,2 structure. Substitution of a sodium ion for the potassium ion in dialkylglycine decarboxylase changes the coordination geometry and results in the movement of the loops that make up the metal ion-binding site (Toney et al., 1993). The dialkylglycine decarboxylase structure provides a precedent for the possibility that the interdomain orientation could be altered in the NgFn1,2 structure by substituting an alkali metal with a different metal to ligand distance for the sodium ion, since the relative orientation of the two Fn-III motifs is partially dictated by the sodium ion at the domain interface.

The discovery of a metal-binding site at the NgFn1,2 domain interface raises the possibility that cation binding is a general feature of Fn-III domain interfaces and might be necessary for the functional stabilization of long filaments of Fn-III domains, such as the five Fn-III repeats of neuroglian. Alternatively, the NgFn1,2 pair may be the only one stabilized by a metal-binding site at its interface, a hypothesis consistent with this fragment's resistance to proteolysis and which may indicate that the first two Fn-III repeats play a unique

role in contact with a receptor. The division of functions in the L1 Fn-III domains, such that the first two promote neurite outgrowth and the last three cause adherence of small cerebellar neurons (Appel et al., 1993), implies distinct roles for the two sets of Fn-III repeats, perhaps suggesting the differential necessity for an interdomain metal-binding site. Using an alignment of the neuroglian and L1 primary sequences (Bieber et al., 1989), we find that three of the four metal ligand residues are conserved (including both of the side chain ligands), suggesting that L1 also contains a metal-binding site at the interface between its first and second Fn-III repeats. The ligand residues are not conserved in the comparable positions at the interfaces between the other domains of neuroglian or L1. Discovery of the metal ion in the NgFn1,2 structure suggests a series of experiments to probe the significance of metals in the function of neuroglian and other cell adhesion molecules. For example, site-directed mutagenesis to change Ser-679 and Asn-743 of neuroglian and the comparable residues of L1 should eliminate the site, and the effects of these replacements could be explored in terms of the protein structure and function, specifically in the ability of these domains to promote neurite outgrowth.

Several lines of evidence suggest that the relative orientations of tandem Fn-III repeats are important for their function in ECM proteins and cell adhesion molecules. For example, tandem presentation of the two Fn-III motifs from N-CAM promotes neurite outgrowth and cell spreading more effectively than either Fn-III domain when presented alone (Frei et al., 1992). Also, the binding of the immunoglobulin superfamily member contactin to the Fn-III domains of tenascin is disrupted by the addition of three extra Fn-III motifs that exist in an alternatively spliced isoform (Zisch et al., 1992). Thus, interactions of receptors with Fn-III motifs are not likely to be confined to a single Fn-III domain, and interactions with adjacent domains can be partially responsible for the ligand specificity of receptors that recognize Fn-III repeats. In other experiments, a proteolytic fragment consisting of Fn-III repeats 8 through 10 of fibronectin was far more effective at promoting adhesion than the FnFn10 domain alone (Nagai et al., 1991). For ECM proteins, it is becoming increasingly clear that neither the RGD sequence alone, nor the domain that contains it, can account for the full cell adhesion properties of ECM proteins containing Fn-III motifs (Yamada, 1991). Thus, contact with receptors must involve additional sites, probably on domains adjacent to the RGD-containing domain.

Recently, two distinct regions in the eighth and ninth Fn-III modules were mapped as synergistic sites that contribute to adhesion properties together with the RGD loop of FnFn10 (Aota et al., 1991; Nagai et al., 1991). Based on the nuclear magnetic resonance structure of FnFn10, the synergistic site on the ninth repeat was suggested to include the C to C' loop (Main et al., 1992). In the NgFn1,2 structure, the C to C' loop

of NgFn1 is ~ 20 Å from the F to G loop of NgFn2. The two loops do not interact directly, but both are on the same face of the molecule and could conceivably be contacted simultaneously. A protein contacting these loops would also be predicted to interact with the B to C loop and the beginning of strand C of the second domain, making this region a potential target for site-directed mutagenesis to define the integrin contact area further.

The orientation between the two neuroglian Fn-III domains is nearly an exact 2-fold screw axis, which was predicted to relate successive Fn-III repeats in tenascin (Leahy et al., 1992). The actual relationship is a 175° rotation followed by a 37.7 Å translation along the rotation axis, which is approximately parallel to the longest dimension of the fragment. Knowing the rotation and translation that relate the two domains in NgFn1,2 allows us to model all five neuroglian Fn-III repeats through the placement of additional domains using the same mathematical operation. The generated five Fn-III repeat structure is approximately a straight rod, a consequence of the near dyad rotation axis and the translation operation being parallel to the longest dimension of the fragment. The model generated in Figure 7 predicts that the Fn-III repeat portion of neuroglian is long (~ 175 Å) and narrow (width, 20–30 Å). Thus far, Fn-III modules in cell adhesion molecules have been found C-terminal to varying numbers of immunoglobulin-related repeats. Information from structural studies of CD4 (Ryu et al., 1990; Wang et al., 1990; Brady et al., 1993) and CD2 (Jones et al., 1992), which respectively consist of four and two tandem immunoglobulin-like repeats, can be used to estimate the approximate dimensions of two tandem immunoglobulin-like repeats as 60–65 Å long and 25–35 Å wide. Combining this information with the model of tandem Fn-III modules presented here predicts that the whole neuroglian molecule (and also its vertebrate homolog L1), with six immunoglobulin-related and five Fn-III domains, could extend up to 370 Å from the cell, assuming no significant bends or kinks between successive domains. Varying the number of immunoglobulin-related repeats and Fn-III modules would cause cell adhesion molecules to either gain or lose length. For example, *Drosophila* fasciclin II (Harrelson and Goodman, 1988) and its vertebrate homolog N-CAM each have five immunoglobulin-like repeats and two Fn-III modules, with a predicted maximum length of ~ 230 Å from the cell surface. Other cell surface molecules contain even greater numbers of tandem Fn-III repeats, such as a human receptor protein tyrosine phosphatase, whose extracellular region consists of 16 repeats (Krueger et al., 1990). In addition to or perhaps alternative to mediating specific interactions with receptors, some of the Fn-III motifs in molecules containing many domains in series may serve as spacer regions to increase the length of the molecule and the distance of functional regions from the cell surface.

Experimental Procedures

Expression

To generate a plasmid for the expression of a soluble form of the entire neuroglian extracellular region, a 3375 bp SacII–AvaI cDNA fragment encoding amino acids 1–1121 (Bieber et al., 1989) was ligated into the KpnI site of the *Drosophila* expression vector pRmHa-3 (Bunch et al., 1988) using KpnI–SacII (GGTACCA-ATTCCGCGG) and AvaI–KpnI (GGACCATGAGCGTGGTACC) linkers. The AvaI–KpnI linker adds a termination codon in the open reading frame, and the resulting plasmid directs the secretion of a 160 kd neuroglian fragment that contains most of the extracellular domain of the molecule, terminating with the proline at amino acid residue 1121.

This neuroglian secretion plasmid was used to create a second plasmid for secretion of the Fn-III repeats. A 1580 bp MluI–KpnI fragment, containing the Fn-III repeats, was excised from the secretion vector and joined to a 118 bp KpnI–HinfI fragment containing the neuroglian leader peptide, using a HinfI–MluI linker (AGAATCGCGCGACG). The resulting plasmid directs the secretion of a neuroglian fragment that contains the five Fn-III repeats and corresponds to amino acids 1–31, 597–1121 (Bieber et al., 1989).

The neuroglian secretion vectors were introduced in *Drosophila* S2 cells (Schneider, 1972) by DNA–calcium phosphate coprecipitation (Wigler et al., 1979). The transformation and selection procedures were essentially as described by Snow et al. (1989) using the plasmid pPC4 (Joker et al., 1989) to confer α -amanitin resistance as a selectable marker. Populations of transformed cells were cloned in soft agar (Cherbas and Cherbas, 1989). Neuroglian secretion from individual cloned lines was induced by addition of cupric sulfate to 0.7 mM, and the lines were tested for high levels of neuroglian secretion by Western blot analyses of culture medium aliquots. Cell lines secreting the entire extracellular domain of neuroglian yielded up to 4 mg of neuroglian protein per liter of culture medium.

Purification

The whole neuroglian extracellular protein and the secreted five Fn-III repeat protein were purified from cell growth medium by immunoaffinity chromatography using the monoclonal antibody 1B7 (Bieber et al., 1989). Eluted protein was further concentrated by vacuum dialysis (Schleicher & Schuell, Inc.) and dialyzed into 100 mM Tris–HCl (pH 8.0), 100 mM NaCl, 2 mM EDTA. Protein concentrations were determined using a BCA assay (Pierce Biochemicals).

Proteolysis to Generate Fragment

Chymotrypsin–Na-p-tosyl-L-lysine or trypsin–N-tosyl-L-phenylalanine chloromethyl (Worthington Biochemicals) was added to the five Fn-III repeat protein (final protease–protein mass ratio of 1:50) in 100 mM Tris–HCl (pH 8.0), 50 mM NaCl, 22 mM CaCl₂. The mixture was incubated at 37°C for 1 hr, and the digestion was stopped by adding phenylmethanesulfonyl fluoride to a final concentration of 1 mM. The major proteolytic fragment was purified by running the digest over a HiLoad 26/60 Superdex 200 Preparative Grade gel filtration column (Pharmacia), and fractions were analyzed by 15% SDS–polyacrylamide gel electrophoresis. Fractions containing the fragment (apparent molecular mass, 37.5 kd) were concentrated with a Centricon 10 microconcentrator (Amicon) to a final concentration of 5 mg/ml in 10 mM Tris–HCl (pH 8.0), 0.05% NaN₃.

N-Terminal Sequencing and Estimation of Fragment Molecular Weight

The purified chymotrypsin and trypsin fragments of the neuroglian Fn-III repeats were analyzed by automated Edman degradation using an Applied Biosystems Model 477A protein sequencer. The chymotrypsin fragment was deglycosylated with PNGase F (Boehringer Mannheim) according to the manufacturer's instructions or with TFMS as described (Edge et al., 1981). Mass spectrometry was performed on a Finnigan Lasermat mass spectrometer.

Crystallization and Preparation of Derivative

Crystallization conditions were initially obtained using a factorial trial screen (Jancarik and Kim, 1991). Crystals were grown by the hanging drop vapor diffusion method (McPherson, 1976) over a reservoir of 1.6 M $(\text{NH}_4)_2\text{SO}_4$, 150 mM Li_2SO_4 , 100 mM HEPES (pH 7.5). The space group of the crystals was identified as F432 ($a = b = c = 241.8 \text{ \AA}$) by screened precession photography. Based on average volume to mass ratios of protein crystals (Matthews, 1968) and assuming a molecular weight of 26 kD for the fragment, the asymmetric unit of the crystal is estimated to contain one fragment and 78% solvent. Crystals were harvested and stored in 100 mM HEPES (pH 7.5), 45% saturated Li_2SO_4 , as no heavy atom compounds were observed to bind to the crystals in an $(\text{NH}_4)_2\text{SO}_4$ harvest buffer (assayed by comparison of screened precession photographs of potential derivatives to native precession photographs). The derivative crystal was soaked overnight in Li_2SO_4 harvest buffer including 10 mM ethylmercuric phosphate.

Cryopreservation of Crystals

Native and derivative crystals were transferred to Li_2SO_4 harvest buffer including 24% glycerol as a cryoprotectant. A single crystal was mounted in a thin film of cryopreservation buffer, supported by a 1 mm diameter glass loop (Teng, 1990), and quickly cooled to -165°C in a nitrogen gas stream using a Siemens LT-2A cryostat. Frozen crystals were screened for mosaic spread and diffraction limit on a Siemens proportional multiwire area detector mounted on a Siemens rotating anode (200 μm focal cup; 50 kV, 90 mA). Reflections were observed from native crystals to $\sim 2.7 \text{ \AA}$ resolution. Acceptable crystals were transported in a liquid nitrogen dewar to CHESS for further data collection.

Collection of Diffraction Data

Data were collected from cryopreserved crystals at -185°C at the CHESS F1 beam line ($\lambda = 0.91 \text{ \AA}$) using an oscillation camera and low temperature device. Diffraction data were recorded on Fuji HR-III image plates and digitized with a BAS 2000 Fuji scanner. Native data (2.0° oscillations) and derivative data (3.0° oscillations) were each collected from a single crystal, with typical exposure time per image of 90–120 s. The relatively long exposure times allowed the collection of native data to $\sim 1.8 \text{ \AA}$ and were not detrimental to the crystals because of the virtual elimination of observed radiation decay at these temperatures (Hope, 1990). This represents an extension of nearly 1 \AA in resolution over the data collectable from the same frozen crystal using the lower flux of X-rays available from a rotating anode.

Data Processing and Reduction

The orientation of the randomly aligned crystal for each data set was determined using the autoindexing routine from the program REFIL (Kabsch, 1988). The orientation was refined, and the raw data were indexed and integrated with the program DENZO, written by Z. Otwinowski. Profile fit intensities were scaled and merged into unique reflections with the ROTAVATA and AGROVATA programs from the CCP4 package (The SERC Collaborative Computing Project No. 4. A Suite of Programs for Protein Crystallography distributed from Daresbury Laboratory, Warrington, WA4 4AD, England [1979]) (Table 1).

Initial Phase Determination

The structure was solved using the isomorphous replacement method with a single heavy atom derivative (single isomorphous replacement). Difference Patterson syntheses were calculated in space group Pm. Two heavy atom sites were located using a general Patterson search procedure (written by B. Hsu). Difference Fourier syntheses revealed the presence of a third lower occupancy site. Phases were calculated including anomalous scattering information (SIRAS phases) in the standard way and refined using the program MLPHARE in the CCP4 package. Despite the low figure of merit and phasing power (Table 2), an initial map calculated at 2.6 \AA showed continuous density and a clear boundary between protein and solvent. The phases were

improved using the solvent-flattening procedure (Wang, 1985) implemented in reciprocal space in the CCP4 package (Leslie, 1987).

Electron Density Interpretation, Refinement, and Structure Analysis

Electron density fitting was performed using the program TOM on IRIS 4D series work stations (Silicon Graphics Incorporated). The program ABONES (Jones and Thirup, 1986) was used to produce a skeleton of the electron density, which clearly showed the positions of two domains that contained the expected Fn-III topology. The carbon- α coordinates of the TnFn3 structure (Leahy et al., 1992) were placed in the electron density using the skeleton, and these coordinates were used as a rough guide during building. The initial model included all main chain residues from Ile-610 (the N-terminus of the chymotrypsin fragment) to Arg-814 (numbered according to Bieber et al., 1989).

All crystallographic refinement, interdomain contact, and solvent accessibility (probe radius, 1.4 \AA) calculations were done using the program X-PLOR (Brünger, 1990). No standard deviation cutoff was applied to the data, and a randomly selected 10% of native structure factor amplitudes were excluded from automated refinement and used to compute a "free" R factor (R_{free}) (Brünger, 1992) throughout refinement. A conventional crystallographic R factor (R_{cryst}) was computed for the remaining reflections that were included in the refinement.

The initial model was refined by conjugate gradient minimization using data between 5 and 3 \AA , followed by additional minimization using data to 2.7 \AA . Several more cycles of manual rebuilding, conjugate gradient minimization, and restrained individual B factor refinement yielded a model ($R_{\text{free}} = 30.4\%$; $R_{\text{cryst}} = 25.6\%$) for data between 5 and 2.5 \AA . Solvent molecules were then placed in peaks that were at least 4 times the rms deviation (σ_{map}) in $(F_{\text{obs}} - F_{\text{calc}})/\Phi_{\text{calc}}$ difference electron density maps. Water molecules were only placed in positions where they could form hydrogen bonds and where it was unlikely that the density corresponded to an alternative side chain position. At $5\text{--}2.3 \text{ \AA}$ ($R_{\text{free}} = 27.7\%$; $R_{\text{cryst}} = 23.2\%$), a sulfate molecule was identified at His-734. At $5\text{--}2.1 \text{ \AA}$ ($R_{\text{free}} = 26.0\%$; $R_{\text{cryst}} = 22.4\%$), strong spherical density was located within $\sim 2.2 \text{ \AA}$ of five potential oxygen ligands and identified as a metal ion because its coordination geometry and ligand distances differed substantially from those expected for ordered water molecules. To obtain unbiased metal to ligand distances, the metal ion was assigned a charge of zero, and van der Waals interactions were minimized by reducing the Lennard-Jones parameters of the ion to near zero. Upon setting the B factor of the ion to the average B factor of the ligand atoms, the occupancy of the metal was refined assuming it was either calcium, magnesium, or sodium. Refinement of either sodium or magnesium yielded an occupancy ≈ 0.77 , and the coordination geometry and ligand distances were compatible with either of these metals. Further minimization and manual intervention resulted in the current model ($5\text{--}2.0 \text{ \AA}$; $R_{\text{free}} = 23.5\%$; $R_{\text{cryst}} = 20.2\%$), which contains residues 610–814, one sodium ion, three N-acetylglucosamine residues, one sulfate molecule, and 237 ordered water molecules (Table 3). The model rms deviations from ideal geometry are 0.015 \AA for bond lengths and 1.9° for angles. The correctness of the structure was verified by a number of evaluative functions, including calculation of a real-space R factor on a per residue basis (Jones et al., 1991), the use of an unbiased R factor during refinement (R_{free} ; Brünger, 1992), and comparison of the main and side chain conformations in the model to a data base of peptide fragments from well-refined structures (Jones et al., 1991).

To characterize the metal-binding site further, data were collected from crystals soaked in Na_2EDTA , CaCl_2 , or NaCl overnight at 22°C . The CaCl_2 soak was preceded by an overnight soak in Na_2EDTA . The Na_2EDTA and CaCl_2 soak solutions consisted of 100 mM HEPES (pH 7.5), 45% saturated Li_2SO_4 , and either 20 mM Na_2EDTA or 10 mM CaCl_2 . The NaCl soak solution consisted of 1.88 M $(\text{NH}_4)_2\text{SO}_4$, 100 mM HEPES/ NH_4OH pH 7.6, 10 mM NaCl . After soaking, crystals were transferred to a buffer composed

of the final soak solution with 25% glycerol as a cryoprotectant. The crystals were flash cooled as described earlier, and data were collected on an R-Axis IIC X-ray detector mounted on a rotating anode.

Difference Fourier analyses revealed a positive electron density peak at the location of the metal ion for all three soak conditions (Na_2EDTA , $4.8\sigma_{\text{map}}$; NaCl , $4.4\sigma_{\text{map}}$; CaCl_2 , $2.7\sigma_{\text{map}}$), demonstrating that the occupancy of the metal ion increased. Assuming it was a sodium ion, the metal occupancy was refined using the 2.0 \AA native coordinates against the structure factors from the three soak data sets, resulting in occupancies of 1.00 (Na_2EDTA data set; $10\text{--}2.65 \text{ \AA}$), 1.04 (NaCl data set; $10\text{--}2.8 \text{ \AA}$), and 0.95 (CaCl_2 data set; $10\text{--}2.65 \text{ \AA}$). The increase in ion occupancy produced by soaking crystals in Na_2EDTA suggests that the ion is sodium rather than magnesium. The NaCl soak also brought the metal ion to full occupancy. The low sodium occupancy observed in the native data set presumably resulted from an exchange of lithium for sodium driven by the high concentration of Li_2SO_4 in the original harvest buffer. The increased occupancy produced by the CaCl_2 soak was likely the result of soaking the crystals in Na_2EDTA prior to the CaCl_2 soak, since the occupancy increase was too low to have been produced by calcium.

Acknowledgments

We thank Michael Blum for help with freezing crystals; Michael Blum and the staff at CHESS for assistance with data collection; Alfonso Mondragón for programs and advice during data processing; Barbara Hsu for her Patterson search program; Art Chirino and Bill Weis for advice during refinement; Dan Leahy, Wayne Hendrickson, and Harold Erickson for the TnFn3 coordinates and many helpful discussions; Abraham De Vos for the human growth hormone receptor coordinates; Bill Lane and the Harvard Microchemical Facility for mass spec and other analyses; the Caltech Microchemical Facility for N-terminal sequence analyses; Peer Bork for aligned sequences of Fn-III repeats; Roland Strong for making Figure 3B; Douglas Rees for helpful discussions about metal sites; and our colleagues for critical reading of the manuscript. The program X-PLOR was run on a CRAY-YMP at the San Diego Supercomputer Center, supported by the National Science Foundation. This work was supported by the Howard Hughes Medical Institute (P. J. B.), a Howard Hughes Medical Institute predoctoral fellowship (A. H. H.), the National Science Foundation (IBN-9120981 to A. J. B.), the American Cancer Society (IRG IN-17 to A. J. B.), and the Purdue Research Foundation (A. J. B.).

Coordinates will be deposited with the Brookhaven Protein Data Bank. Until processing is complete, they will be available by e-mail (ahhacitray.caltech.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received November 4, 1994; revised January 21, 1994.

References

- Adzhubei, A. A., and Sternberg, M. J. E. (1993). Left-handed polypeptide II helices commonly occur in globular proteins. *J. Mol. Biol.* 229, 472–493.
- Aota, S., Nagai, T., and Yamada, K. M. (1991). Characterization of regions of fibronectin besides the arginine-glycine-aspartic acid sequence required for adhesive function of the cell-binding domain using site-directed mutagenesis. *J. Biol. Chem.* 266, 15938–15943.
- Appel, F., Holm, J., Conscience, J.-F., and Schachner, M. (1993). Several extracellular domains of the neural cell adhesion molecule L1 are involved in neurite outgrowth and cell body adhesion. *J. Neurosci.* 13, 4764–4775.
- Baron, M., Main, A. L., Driscoll, P. C., Mardon, H. L., Boyd, J., and Campbell, I. A. (1992). ^1H NMR assignment and secondary structure of the cell adhesion type III module of fibronectin. *Biochemistry* 31, 2068–2073.
- Bazan, J. F. (1990). Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. Natl. Acad. Sci. USA* 87, 6934–6938.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542.
- Bieber, A. J., Snow, P. M., Hortsch, M., Patel, N. H., Jacobs, J. R., Traquina, Z. R., Schilling, J., and Goodman, C. S. (1989). Drosophila neuroglian: a member of the immunoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. *Cell* 59, 447–460.
- Bork, P., and Doolittle, R. F. (1992). Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl. Acad. Sci. USA* 89, 8990–8994.
- Brady, R. L., Dodson, E. J., Dodson, G. G., Lange, G., Davis, S. J., Williams, A. F., and Barclay, A. N. (1993). Crystal structure of domains 3 and 4 of rat CD4: relation to the NH_2 -terminal domains. *Science* 260, 979–983.
- Brünger, A. T. (1990). X-PLOR (Version 2.1) Manual (New Haven, Connecticut: Yale University).
- Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
- Bunch, T. A., Grinblatt, Y., and Goldstein, L. S. B. (1988). Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells. *Nucl. Acids Res.* 16, 1043–1061.
- Cherbas, L., and Cherbas, P. (1989). Cloning tissue culture cells. In *Drosophila: A Laboratory Manual*, M. Ashburner, ed. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 136–137.
- De Vos, A. M., Ultsch, M., and Kossiakoff, A. A. (1992). Human Growth Hormone and Extracellular Domain of Its Receptor: Crystal Structure of the Complex. *Science* 255, 306–312.
- Doolittle, R. F. (1987). Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences (Mill Valley, California: University Science Books).
- Edge, A. S. B., Faltynek, C. R., Hof, L., Reichert, L. E., Jr., and Weber, P. (1981). Deglycosylation of glycoproteins by trifluoromethanesulfonic acid. *Anal. Biochem.* 118, 131–137.
- Evans, S. V. (1993). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* 4, 134–138.
- Frei, T., von Bohlen und Halbach, F., Wille, W., and Schachner, M. (1992). Different extracellular domains of the neural cell adhesion molecule (N-CAM) are involved in different functions. *J. Cell Biol.* 118, 177–194.
- Glusker, J. P. (1991). Structural requirements of metal liganding to functional groups in proteins. *Adv. Prot. Chem.* 42, 1–76.
- Grenningloh, G., Bieber, A. J., Rehm, E. J., Snow, P. M., Traquina, Z. R., Hortsch, M., Patel, N. H., and Goodman, C. S. (1990). Molecular genetics of neuronal recognition in *Drosophila*: evolution and function of the immunoglobulin superfamily cell adhesion molecules. *Cold Spring Harbor Symp. Quant. Biol.* 55, 327–340.
- Harrelson, A. L., and Goodman, C. S. (1988). Growth cone guidance in insects: fasciclin II is a member of the immunoglobulin superfamily. *Science* 242, 700–708.
- Holmgren, A., and Bränden, C.-I. (1989). Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature* 342, 248–251.
- Hope, H. (1990). Crystallography of biological macromolecules at ultra-low temperature. *Annu. Rev. Biophys. Chem.* 19, 107–126.
- Hortsch, M., and Goodman, C. S. (1991). Cell and substrate adhesion molecules in *Drosophila*. *Annu. Rev. Cell Biol.* 7, 505–557.

- Hynes, R. O. (1990). *Fibronectins* (New York: Springer-Verlag).
- Jancarik, J., and Kim, S. H. (1991). Sparse-matrix sampling—a screening method for crystallization of proteins. *J. Appl. Crystallogr.* 24, 409–411.
- Jokerst, R. S., Weeks, J. R., Zehring, W. A., and Greenleaf, A. L. (1989). Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. *Mol. Gen. Genet.* 215, 266–275.
- Jones, T. A., and Thirup, S. (1986). Using known substructures in protein model building and crystallography. *EMBO J.* 5, 819–822.
- Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119.
- Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K., and Stuart, D. I. (1992). Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature* 360, 232–239.
- Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr.* 21, 67–71.
- Krueger, N. X., Streuli, M., and Saito, H. (1990). Structural diversity and evolution of human receptor-like protein tyrosine phosphatases. *EMBO J.* 9, 3241–3252.
- Kwong, P. D., Ryu, S. E., Hendrickson, W. A., and Axel, R. (1990). Molecular characteristics of recombinant human CD4 as deduced from polymorphic crystals. *Proc. Natl. Acad. Sci. USA* 87, 6423–6427.
- Leahy, D. J., Hendrickson, W. A., Aukhil, I., and Erickson, H. P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* 258, 987–991.
- Leslie, A. G. W. (1987). A reciprocal space method for calculating a molecular envelope using the algorithm of Wang, B. C. *Acta Crystallogr. A* 43, 134–136.
- Main, A. L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. A. (1992). The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* 71, 671–678.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* 33, 491–497.
- McPherson, A. (1976). The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. In *Methods of Biochemical Analysis*, D. Glick, ed. (New York: John Wiley & Sons, Inc.), pp. 249–345.
- Nagai, T., Yamakawa, N., Aota, S., Yamada, S. S., Akiyama, S. K., Olden, K., and Yamada, K. M. (1991). Monoclonal antibody characterization of two distant sites required for function of the central cell-binding domain of fibronectin in cell adhesion, cell migration, and matrix assembly. *J. Cell Biol.* 114, 1295–1305.
- Patthy, L. (1990). Homology of a domain of the growth hormone/prolactin receptor family with type III modules of fibronectin. *Cell* 67, 13–14.
- Pierschbacher, M. D., and Ruoslahti, E. (1984). Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature* 309, 30–33.
- Richardson, J. S., Getzoff, E. D., and Richardson, D. G. (1978). The β bulge: a common small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci. USA* 75, 2574–2578.
- Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R. W., and Hendrickson, W. A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* 348, 419–426.
- Schneider, I. (1972). Cell lines derived from the late embryonic stages of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* 27, 353–365.
- Snow, P. M., Bieber, A. J., and Goodman, C. S. (1989). Fasciclin III: a novel homophilic adhesion molecule in *Drosophila*. *Cell* 59, 313–323.
- Teng, T.-Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Crystallogr.* 23, 387–391.
- Toney, M. D., Hohenester, E., Cowan, S. W., and Jansonius, J. N. (1993). Dialkylglycine decarboxylase structure: bifunctional active site and alkali metal sites. *Science* 261, 756–759.
- Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Meth. Enzymol.* 90, 90–112.
- Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* 348, 411–419.
- Wigler, M., Pellicer, A., Silverstein, S., Avel, R., Urlaub, G., and Chasin, L. (1979). DNA-mediated transfer of the adenine phosphoribosyltransferase locus into mammalian cells. *Proc. Natl. Acad. Sci. USA* 76, 1373–1376.
- Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily—domains for cell surface recognition. *Annu. Rev. Immunol.* 6, 381–405.
- Yamada, K. M. (1991). Adhesive recognition sequences. *J. Biol. Chem.* 266, 12809–12812.
- Yamashita, M. M., Wesson, L., Eisenman, G., and Eisenberg, D. (1990). Where metal ions bind in proteins. *Proc. Natl. Acad. Sci. USA* 87, 5648–5652.
- Yoshihara, Y., Oka, S., Ikeda, J., and Mori, K. (1991). Immunoglobulin superfamily molecules in the nervous system. *Neurosci. Res.* 10, 83–105.
- Zisch, A. H., D'Alessandri, L., Ranscht, B., Falchetto, R., Winterhalter, K. H., and Vaughan, L. (1992). Neuronal cell adhesion molecule contactin/F11 binds to tenascin via its immunoglobulin-like domains. *J. Cell Biol.* 119, 203–213.

Chapter 3

Refinement of the Structure: A Detailed Description

Introduction

This chapter describes the refinement of the crystallographic structure of the first two type III fibronectin repeats from *Drosophila* neuroglian (NgFn1,2). The goal of crystallographic refinement is to produce a structural model consistent with the observed diffraction data and standard stereochemical constraints. The overall refinement process is iterative and the production of the final model generally requires several refinement “cycles” or “stages,” each of which can be further divided into three distinct steps: model building, automated refinement and evaluation of the refined model.

Model Building

Crystallographic refinement starts as soon as an initial model of the structure has been completed. Unfortunately, the experimentally derived phases for a structure are often inaccurate and to relatively low resolution (3.5-3.0Å), resulting in electron density maps that are ambiguous and difficult to trace. Thus, the models used during the early stages of refinement are commonly incomplete and have poor geometry.

After completion of the first cycle of refinement, the refined model is modified and hopefully improved using the initial maps as well as difference maps as a guide. During the first few cycles of refinement, the new maps are generally calculated using the experimental phases or some combination of experimental and calculated phases (ϕ_{calc} , phases calculated from the model). The most commonly used difference maps are calculated with $[2F_o - F_c]$ or $[F_o - F_c]$ Fourier coefficients, where F_o and F_c respectively represent observed structure factors and structure factors calculated from the model. Errors in the structure are highlighted in $[F_o - F_c]$ maps, with positive electron density peaks corresponding to missing atoms and negative density corresponding to included atoms that are incorrectly placed. The interpretation of these maps is sometimes difficult because they

are often quite noisy, and correctly placed atoms with artificially low B-values will generate negative difference density. Maps calculated with [2Fo-Fc] Fourier coefficients are equivalent to the summation of maps calculated with Fo and [Fo-Fc] coefficients. These maps also highlight errors in the model but do so in the context of the rest of the molecule's electron density. Although the maps generated during refinement are generally of much better quality than the initial maps, they are biased by the refinement process and sometimes contain errors. If the experimental phases were generated using isomorphous replacement or MAD phasing, the initial maps are unbiased and can periodically be used to check the validity of the refined model.

As the refinement progresses, the phases calculated from the refined model become the best estimate of the true phases. Eventually, the new maps are generated using only calculated phases. This switch from experimental to calculated phases enables phase extension beyond the resolution achieved with the derivative data.

Automated Refinement

Differences between observed and calculated structure factors as well as errors in model stereochemistry are minimized using some form of automated refinement. This process is limited during the early stages of a structure refinement by the low resolution of the initial phase information. Each atom in a model is described by at least five parameters: the x, y, z coordinates of the atom, an isotropic temperature factor B (\AA^2), which describes the thermal motion of the atom, and Q , the occupancy of the atomic position. The first four of these parameters are commonly refined for all atoms in a model, leading to a minimization calculation with several thousand parameters. At the resolutions commonly used during the early stages of the refinement process, the low number of possible data points (structure factors) frequently results in a minimization problem that is only slightly overdetermined or sometimes even underdetermined. This initial paucity of observed data is partially

compensated for by including stereochemical restraints in the refinement process. Eventually, the refined phases are extended, allowing the use of additional (higher resolution) native data.

X-PLOR is a program commonly used for macromolecular refinement and molecular dynamics¹. The program refines an atomic model by minimizing the value of E_{TOTAL} , an energy function that describes the overall energy for the input structure. E_{TOTAL} is actually the sum of two different “classes” of calculated energy: $E_{EMPIRICAL}$, which is derived from model geometry and crystal packing, and $E_{EFFECTIVE}$, which is commonly related to the correlation between calculated and observed data.

$$E_{TOTAL} = E_{EMPIRICAL} + E_{EFFECTIVE} \quad (3-1)$$

$E_{EMPIRICAL}$ includes several independent energy terms that are functions of model geometry and nonbonded interactions. The value of each term is calculated using force constants empirically derived from small molecule crystal structures and spectroscopy.

$$E_{EMPIRICAL} = E_{BOND} + E_{ANGLE} + E_{DIHE} + E_{IMPR} \quad (3-2)$$

$$+ E_{HBON} + E_{VDW} + E_{ELEC} + E_{PVDW}$$

$$+ E_{PELE}$$

The terms E_{BOND} , E_{ANGLE} , E_{DIHE} and E_{IMPR} describe the energy of atomic interactions that involve covalent bonds. E_{BOND} and E_{ANGLE} are the bond and bond angle energy terms while E_{DIHE} and E_{IMPR} involve dihedral angles, chirality and planarity. E_{HBON} , E_{VDW} and E_{ELEC} describe noncovalent (nonbonded) energies that are a result of intramolecular interactions or intermolecular interactions between molecules related by non-crystallographic symmetry. E_{VDW} and E_{ELEC} are terms for van der Waals and electrostatic energies, respectively. E_{HBON} represents a hydrogen bond energy term that is only used

with certain force fields. van der Waals and electrostatic interactions resulting from the packing of symmetry related molecules are described by $E_{\text{VDW}} + E_{\text{PELE}}$.

$E_{\text{EFFECTIVE}}$ is an “energy” target function composed of several terms, only a few of which are used during any given type (crystallographic, NMR) or cycle of refinement. Some terms are used to restrain or constrain changes in model coordinates or angles and others are used to measure the correlation between a model and the observed data. $E_{\text{X-RAY}}$, a component of $E_{\text{EFFECTIVE}}$, is the crystallographic target function, which measures how well the model explains the observed crystallographic data. X-PLOR provides a choice of several possible target functions for the $E_{\text{X-RAY}}$ term, only one of which can be used during a given minimization job. The most commonly used (default) target for $E_{\text{X-RAY}}$ is based on the crystallographic residual and takes the form

$$E_{\text{X-RAY}} = W \sum_h w_h [|F_o(h)| - k |F_c(h)|]^2 + E_{\text{P}_{\text{X-RAY}}} \quad (3-3)$$

where h represents (h,k,l) , the Miller indices of the selected reflections; W is an overall weight factor for the $E_{\text{X-RAY}}$ term; w_h is an individual weight for each reflection; F_o and F_c are observed and calculated structure factors; k is a scale factor; and $E_{\text{P}_{\text{X-ray}}}$ is a term that can be used to add phase restraints to the refinement.

Two different approaches can be used to minimize E_{TOTAL} within X-PLOR: conjugate gradient minimization using the Powell algorithm² or a molecular dynamics based simulated annealing. Conjugate gradient minimization will effectively minimize the energy target function. However, since this algorithm only moves down a gradient, it is susceptible to getting stuck in local minima; a shortcoming which sometimes makes this algorithm unsuitable for problems with multiple local minima, such as crystallographic refinement.

In *simulated annealing*, the model is heated to a defined “temperature” by assigning initial velocities, from a Maxwellian distribution, to the atoms of the structure. The atoms of the model are then allowed to move freely within the constraints provided by the force field, a process commonly referred to as molecular dynamics. If the kinetic energy of the system is high enough, the structure can escape local minima by temporarily moving up an energy gradient. At regular intervals, the model is “cooled” by slightly reducing the energy of the system, a process which results in a dampening of atomic motion. The molecular dynamics simulations are stopped when the overall kinetic energy of the system has reached a level at which it is unlikely the model will escape the nearest energy minimum. At this point, several cycles of Powell-method conjugant gradient minimization are used to bring the model to the nearest energy minimum and to regularize the stereochemistry.

The refinement of an overall B-factor for the structure and the refinement of atomic occupancies are both carried out using conjugate gradient minimization with EX-RAY as the target function. Individual isotropic atomic B-factors are also minimized by conjugate gradient minimization. In this case the target function is

$$T = E_{X-RAY} + W_B E_R \quad (3-4)$$

where E_{X-RAY} is the crystallographic target function, W_B is a weight factor (included within the term E_R in the X-PLOR manual) and E_R is a restraining term used to limit the deviation in atomic B-values between atoms joined by covalent bonds and bond angles. If W_B is not defined by the user, a value is calculated by X-PLOR (automatic weighting). Unrestrained atomic B refinement can be carried out by setting the value of W_B to zero.

Structural Analysis

At the end of each cycle of refinement, the model’s deviation from ideal stereochemistry and the observed data needs to be evaluated. This is an important step in each cycle that

enables one to assess the effectiveness of the refinement process as well as the reliability of the refined structure. The information obtained by monitoring changes in model reliability as a function of refinement cycle is used to choose the force fields and the method of minimization, to modify weighting terms, and to decide which terms are to be included in $E_{\text{EFFECTIVE}}$, resulting in an overall ‘tuning’ of the refinement process. An analysis of model stereochemistry on a per residue basis is also very useful because it highlights energetically unfavorable regions of the model that are probably incorrect and need to be fixed.

The two most commonly quoted indicators of overall structural reliability are the resolution to which the structure was refined and the crystallographic residual (R-factor). The R-factor, which is sometimes quoted as a percentage, is calculated using the model and the observed diffraction data and is defined as

$$R = \sum_h |F_o - F_c| / \sum_h |F_o| \quad (3-5)$$

where h represents (h,k,l) , the Miller indices of the selected reflections. While the R-factor does give a rough estimate of overall model quality it is generally insensitive to small regions of incorrect structure if the protein is large. The R-factor has the additional disadvantage of being closely related to the crystallographic target function commonly used during energy minimization. This raises the possibility that the R-factor, which is supposed to reflect the information content of the model, can be artificially lowered or “biased” by the energy minimization algorithm. An unbiased or “free” R-factor (R_{free})³ can be calculated using a randomly selected subset (typically 5-10%) of the observed data that is excluded from minimization calculations during the entire refinement process. The conventional crystallographic R-factor (R_{cryst}) is calculated with the data used in the refinement process. The average positional error in model coordinates, another indicator of

model reliability, can be estimated with a Luzzati plot⁴, which is constructed by plotting the R-factor as a function of resolution.

It is also common to evaluate the reliability of a refined model through a detailed examination of model geometry. Some characteristics, such as the model's root mean square (r.m.s.) deviation from ideal bond distances and angles, are commonly quoted to show that a model is stereochemically sound. Another barometer of stereochemical reasonability is the Ramachandran plot. This plot is a graphical representation of the polypeptide backbone dihedral angles, phi and psi. The phi, psi combinations that can be adopted by non-glycine amino acids are highly constrained by unfavorable steric collisions. Residues with disallowed phi, psi combinations are in an energetically unfavorable conformation and are probably incorrectly placed in the model. Thus, a Ramachandran plot of the entire protein is a graphical representation of the overall quality of the structure. These plots are also useful during the building of a model because they show which residues or regions of the structure have unfavorable conformations. Some of the other stereochemical quantities that are reported on a per residue basis and used during model building are side chain dihedral angles, peptide bond planarity and carbon- α chirality.

Completion of Refinement

It can be difficult to tell when to actually stop the refinement process because the model will asymptotically approach the true representation of the structure. Each new set of maps will reveal "new" inconsistencies and errors in the current model. Thus, finishing refinement is essentially deciding that the remaining errors are minor and do not affect the interpretation of the structure. This point is usually reached when difference Fouriers do not reveal any large unexplained peaks and when the structure has been refined to the limits of the available data with a good R-factor (20% or lower) and good geometry (r.m.s. bond $< 0.03\text{\AA}$ and r.m.s. angle $< 3^\circ$).

Materials and Methods

The Initial Structure

As described in Chapter 2, the initial phases for the structure were determined to 2.6 Å with a single isomorphous derivative and were used to calculate an initial 2.6 Å map which was improved by solvent flattening. An electron density “skeleton” calculated from the initial map revealed two domains with the expected topology. The carbon- α coordinates of the TnFn3 (third type III fibronectin repeat from tenascin) structure⁵, placed in the initial map using the skeleton, were used as a rough guide during the building of the initial model. The initial model included all backbone atoms from residue Ile610 (the N-terminus of the chymotrypsin fragment as numbered previously⁶) to Arg814 and three N-acetylglucosamine (NAG) residues. All atoms were given a B-value of 20 Å².

Refinement

All crystallographic refinement of model parameters (atomic coordinates, temperature factors and occupancies) was performed with the program X-PLOR¹. All molecular dynamics and energy minimization calculations were carried out using the protein force field parhcsdx.pro⁷, the carbohydrate force field param3.cho⁸ and the param19.sol⁹ force field for ordered water molecules.

Structure Analysis

The programs X-PLOR¹, PROCHECK¹⁰, and O¹¹ were used to analyze the overall quality of the model, model stereochemistry on a per residue basis, and the real space correlation between the model and the observed electron density, respectively. X-PLOR was used to calculate R_{cryst} and R_{free} as well as r.m.s. deviations in bond lengths and angles. This program was also used to analyze crystal packing and solvent accessibility. A detailed examination of model stereochemistry was performed with PROCHECK, which

compares the geometry of the model to that observed in a database of well refined structures. The results of the analysis, which are output numerically and graphically, are evaluated for the structure as a whole as well as on a per residue basis. The program O was used to calculate a real-space R-factor on a per residue basis for backbone and side chain atoms. The experimental 2.6 Å solvent flattened map and refined [2Fo-Fc] maps were used in the calculations.

Model Building

Model building was carried out with the interactive graphics programs TOM and O¹¹ on IRIS 4D series workstations (Silicon Graphics Incorporated). The building of the initial model was described above. Most of the subsequent models were constructed using four maps: the experimental 2.6 Å SIRAS map (phases derived with Single Isomorphous Replacement + Anomalous Scattering), the 2.6 Å solvent flattened SIRAS map, and the most recent [2Fo-Fc], ϕ_{calc} and [Fo-Fc], ϕ_{calc} difference maps. The difference maps were calculated with a low resolution cutoff of 10 Å because including the low resolution terms appeared to yield cleaner maps with better connectivity.

Results and Discussion

The Refinement Process

The NgFn1,2 model can be viewed as the product of two individual refinement efforts. In the first cycle of the first refinement attempt, the initial model was minimized with simulated annealing using data from 6 to 2.1 Å. High resolution data are not generally included during the early stages of a crystallographic refinement because they decrease the radius of convergence; the distance, in parameter space, the true minimum can be from the starting point and still be found by the minimization algorithm. In this case the high resolution data were included because the high quality of the initial maps allowed a reliable

tracing of the polypeptide backbone and because the model was being refined by simulated annealing. When a high initial temperature is used, simulated annealing can move atoms large distances and appears to have a larger radius of convergence than other algorithms. The results of the first refinement cycle looked quite promising. The R-factors for the initial model ($R_{\text{free}} = 40.80$, $R_{\text{cryst}} = 40.46$) were substantially reduced ($R_{\text{free}} = 34.88$, $R_{\text{cryst}} = 30.52$) and the geometry of the refined model was reasonable (r.m.s. bond and angle deviations of 0.026 Å and 2.50° respectively).

The second and subsequent refinement cycles used conjugant gradient minimization so that the effort spent during the rebuilding process was not lost during molecular dynamics. By the beginning of the fourth refinement cycle, it was becoming difficult to tell how to proceed. The model had reasonable geometry (r.m.s. bond and angle deviations of 0.014 Å and 1.87°, respectively) and very few errors were evident in [Fo-Fc] difference maps. The placement of 62 ordered water molecules improved the R-factors ($R_{\text{free}} = 34.88$, $R_{\text{cryst}} = 30.52$) but still did not reduce R_{cryst} below 25%, the value at which the refinement of incorrect structures frequently stalls¹². At this point, it was decided that the refinement was not proceeding as one would expect for a correct model and that the inclusion of data to 2.1 Å during the initial refinement cycle was a mistake. Refinement cycle 5 was stopped after 15 cycles of minimization because there was little change in the R-factors. In preparation for the second refinement process, individual atomic B-factors were discarded and an overall B-factor was refined for the structure.

The “new” refinement process began with cycle 6, in which the model from the cycle 5 rebuilding step was refined with conjugate gradient minimization using data from 5.0 to 3.0 Å. Several more refinement cycles using conjugate gradient minimization produced the final model. A detailed account of the cycles from both refinement attempts is given below.

Refinement cycle 1. Before minimization, the value of W (equation 3-3) that would weight $E_{\text{EMPIRICAL}}$ and $E_{\text{X-RAY}}$ approximately equally (W_{ideal}) was determined using the CHECK protocol distributed with X-PLOR. W was set to twice the calculated “ideal” value in order to increase the importance of the observed data in the minimization process and thereby avoid placing unnecessarily rigid constraints on atomic motion. Subsequent refinement cycles also used twice the ideal $E_{\text{X-RAY}}$ weight.

Simulated annealing, as implemented in X-PLOR, commonly involves three independent minimizations: “prepstage” (PS), simulated annealing (SA), and conjugate gradient minimization (CG). In the prepstage step, the carbon- α atoms of the input model are restrained to their initial coordinates by a harmonic energy term (part of $E_{\text{EFFECTIVE}}$) and bad contacts are eliminated with conjugate gradient minimization. This step is often necessary because strain or bad contacts present in the initial model can cause the model to “fly-apart” when heated during simulated annealing. The post-prepstage model of NgFn1,2 was minimized by simulated annealing using an initial temperature of 3000, 50 molecular dynamics steps of 0.5 femtoseconds at each temperature, and cooling steps of 25. The simulated annealing output was then subjected to 120 cycles of conjugate gradient minimization.

The first refinement cycle decreased R_{cryst} by 9.9% and R_{free} by 5.9% (Table 3-1), with the largest drop occurring during prepstage. None of the loops that were poorly defined in the original maps had adopted new conformations as a result of the molecular dynamics. However, a comparison of the refined and the input model coordinates revealed that several residues had moved significantly from their initial positions. Some of these new residue positions were obviously inconsistent with the initial electron density maps.

Table 3-1 Statistics for refinement cycles 1 to 5.

Refinement Cycle	Type of Refinement	Resolution (Å)	INPUT MODEL					OUTPUT MODEL				
			A. A.	Sugars	Waters	Other	R _{free}	R _{cryst}	rms Bond	rms Angle	R _{free}	R _{cryst}
1	40PS	6.0 - 2.1	205	3	0	0	40.80	40.46			36.5	32.7
	SA, 3000	6.0 - 2.1									34.9	30.6
	120CG	6.0 - 2.1							0.026	2.50	34.88	30.52
2	25CG, 20IB	6.0 - 2.1	200	1	0	0	37.50	36.34	0.017	2.06	34.15	31.96
3	25CG, 20AB	6.0 - 2.1	202	1	0	0	33.98	31.88	0.014	1.87	32.04	29.19
4	25CG, 10AB	6.0 - 2.1	202	1	49	0	30.68	28.18	0.012	1.78	29.26	26.43
5	15CG	6.0 - 2.1	204	2	62	0	29.51	26.68			29.52	26.57
	20IB	6.0 - 2.1							0.013	1.87	31.05	28.43

PS = cycles of Powell method conjugate gradient minimization in "prestage"; CG = cycles of Powell method conjugate gradient minimization; AB = cycles of minimization of restrained, individual, isotropic atomic temperature factors; UAB = cycles of minimization of unrestrained, individual, isotropic atomic temperature factors; IB = cycles of minimization of an overall isotropic B factor for the structure; A.A. = amino acid residues; sugars = N-acetylglucosamine residues; waters = ordered water molecules; S = sulfate anion; M = magnesium cation; N = sodium cation; r.m.s. bond = root mean square deviation from ideal bond lengths; r.m.s. angle = root mean square deviation from ideal bond angles; R_{cryst} = $\Sigma_a (|F_o - F_c|) / (\Sigma_a F_o)$, where F_o and F_c are observed and calculated structure factor amplitudes and a_e (refinement-set reflections); R_{free} = $\Sigma_t (|F_o - F_c|) / (\Sigma_t F_o)$, where t ∈ (free-set reflections);

Refinement cycle 2. Powell-method conjugant gradient minimization was used to refine the model's atomic coordinates (positional refinement) during this and all subsequent refinement cycles (Table 3-1). The rebuilt model did not include Ile610, Arg814 or residues Pro758-Ala760 because their electron density was weak or ambiguous in the ($[2\text{Fo}-\text{Fc}], \phi_{\text{calc}}$) map. Positional refinement was followed by overall B refinement, which optimizes the overall B-value of the model by calculating a shift that is applied to all atomic B-values. Refinement cycle 2 resulted in an R_{cryst} that was higher than that of the previous cycle (Table 3-1). However, the value of R_{free} decreased 0.73%, indicating that the model had actually been improved. The positive shift in R_{cryst} and the relatively small negative shift in R_{free} were not very encouraging. Early in the refinement process, one generally expects to observe substantial drops in the R-factors due to the correction of major errors present in the initial model.

Refinement cycle 3. The $[2\text{Fo}-\text{Fc}], \phi_{\text{calc}}$ map generated after refinement cycle 2 was well defined and continuous except for a small stretch of weak backbone density at residues Pro758 and Ala759. The $[\text{Fo}-\text{Fc}], \phi_{\text{calc}}$ difference map was relatively clean except for density that apparently belonged to ordered water molecules. Since the agreement between the input model (refinement cycle 2 output) and the new difference maps was quite good, despite the relatively poor R-factors, it was difficult to locate errors in the model. The new model contained all residues from Ile610 to Asp813 except for Pro758 and Ala759.

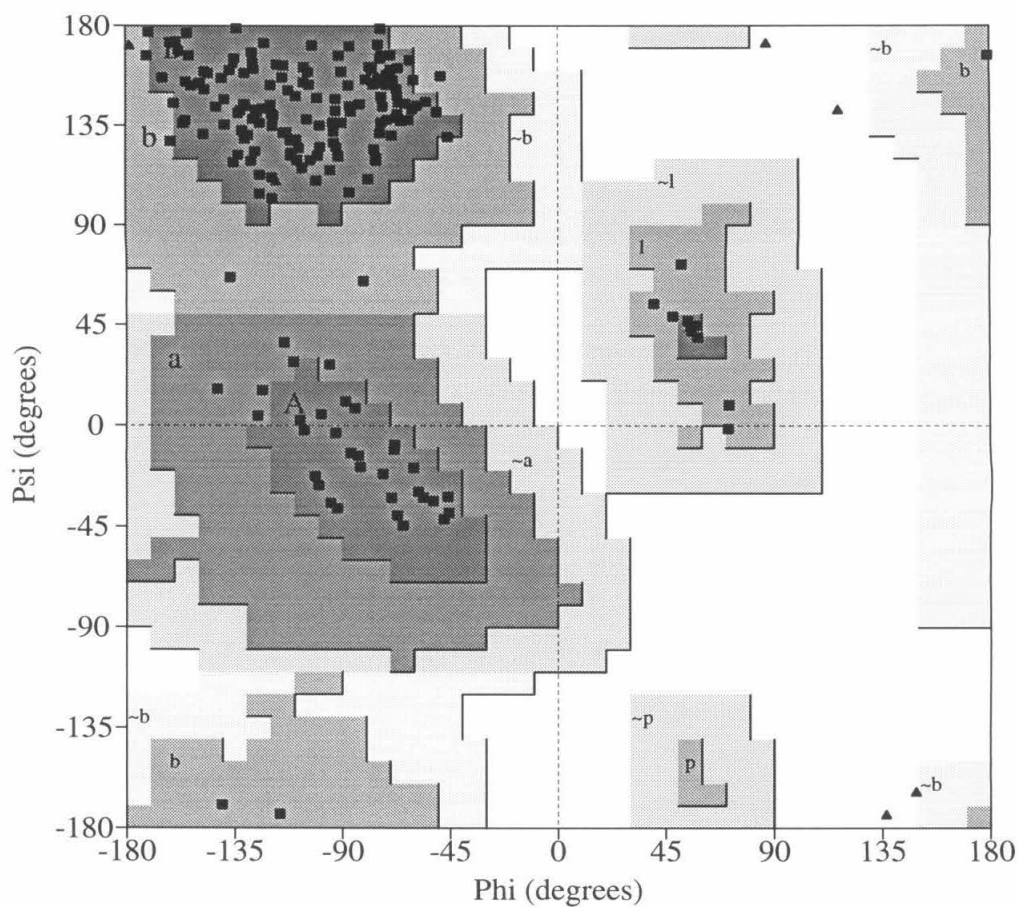
Positional refinement was followed by the refinement of restrained, individual atomic B-values because it was evident that positional refinement alone was not going to significantly reduce the R-factors. The initial refinement of atomic B-values commonly produces a significant drop in R_{cryst} because the additional parameters enable the model to better approximate reality. It was hoped that this step would significantly improve the

phases and thereby make it easier to locate model errors. However, the introduction of additional refinement parameters can also “artificially” lower R_{cryst} and obscure errors in the model. Refinement cycle 3 produced a substantial (2.11%, Table 3-1) drop in R_{free} , indicating that the model had been improved (Table 3-1), but R_{cryst} was still quite high for a model refined to 2.1 Å with good geometry and relatively clean difference maps.

Refinement cycle 4. Once again, both difference maps ($[2\text{Fo}-\text{Fc}], \phi_{\text{calc}}$ and $[\text{Fo}-\text{Fc}], \phi_{\text{calc}}$) indicated that the input model was essentially correct. Several large positive peaks, which appeared to be ordered water molecules, were present in the $[\text{Fo}-\text{Fc}]$ map. The placement of water molecules usually produces substantial drops in R_{cryst} , but can lead to errors in the model that are difficult to detect after minimization. These errors, which usually involve placing water molecules in protein electron density, may produce an initial drop in R_{cryst} but result in higher final R-factors. The best way to avoid such errors is to place the water molecules only after the model has been refined to a reliable value of R_{cryst} . At this point in the refinement of the NgFn1,2 structure, R_{cryst} was still dangerously high but the placement of water molecules was the only correction that clearly needed to be made. Water molecules were placed in $+4\sigma$ $[\text{Fo}-\text{Fc}]$ difference peaks that allowed the formation of chemically reasonable hydrogen bonds. Minimization stalled after 25 cycles and was terminated. The resulting model had tight geometry (r.m.s. bond and angle deviations of 0.012 Å and 1.8° respectively) and improved R-factors ($R_{\text{free}} = 29.26$, $R_{\text{cryst}} = 26.43$) but R_{cryst} was still above 25% and in the “danger zone” proposed by Brändén and Jones¹². The stereochemistry of the model was examined with PROCHECK and found to be reasonable (Figure 3-1).

Figure 3-1 (following pages) Modified Procheck output from analysis of model produced by refinement cycle 4. (A) Ramachandran plot of model phi, psi angles. An analysis of 118 structures solved to at least 2.0 Å resolution with an R_{cryst} of no greater than 20% suggests that a good quality model will have over 90% of its phi, psi angles in the most favored regions. The most favorable regions of phi, psi space have the darkest shading in the diagram. Less favorable regions have less shading. (B) Plot of chi-1 versus chi-2. The dashed crosses represent the ideal gauche and trans regions for chi-1 and chi-2 dihedral angles. The crosses extend to one standard deviation from the ideal values and points that are more than 2.5 standard deviations from the ideal are shown as empty squares. This plot includes 107 of the 202 residues present in the model. Fourteen of these residues are represented as empty squares. This program artificially inflates the number of residues with poor chi-1, chi-2 angles by including asparagine residues in the analysis. These residues do not have ideal chi-2 values at 60, 180 and 300°. (C) Assessment of quality of polypeptide backbone parameters. These plots present the overall quality of the polypeptide backbone's stereochemistry in terms of the following parameters: phi and psi dihedral angles, peptide bond planarity, bad non-bonded interactions, carbon- α tetrahedral distortion, and hydrogen bond energies. The line through the middle of the shaded area in each plot is the mean value expected for a well refined structure. The shaded area itself represents one standard deviation about the mean. The value for the NgFn1,2 structure is represented by a shaded square. (D) Assessment of the quality of side chain parameters. These plots present the overall reliability of the placement of side chains in the model in terms of chi-1 and chi-2 angles. The shaded band in each plot represents one standard deviation about the mean value (central line) expected for a well refined model. The value for the NgFn1,2 model is represented by a shaded square. (E, F and G) Plot of stereochemical parameters on a per residue basis. Deviations in chi-1 angles, peptide bond planarity and carbon- α chirality are plotted on a per residue basis allowing the identification of problematic areas in the model. Highlighted residues are those that deviate by more than 2.0 standard deviations from ideal.

A.

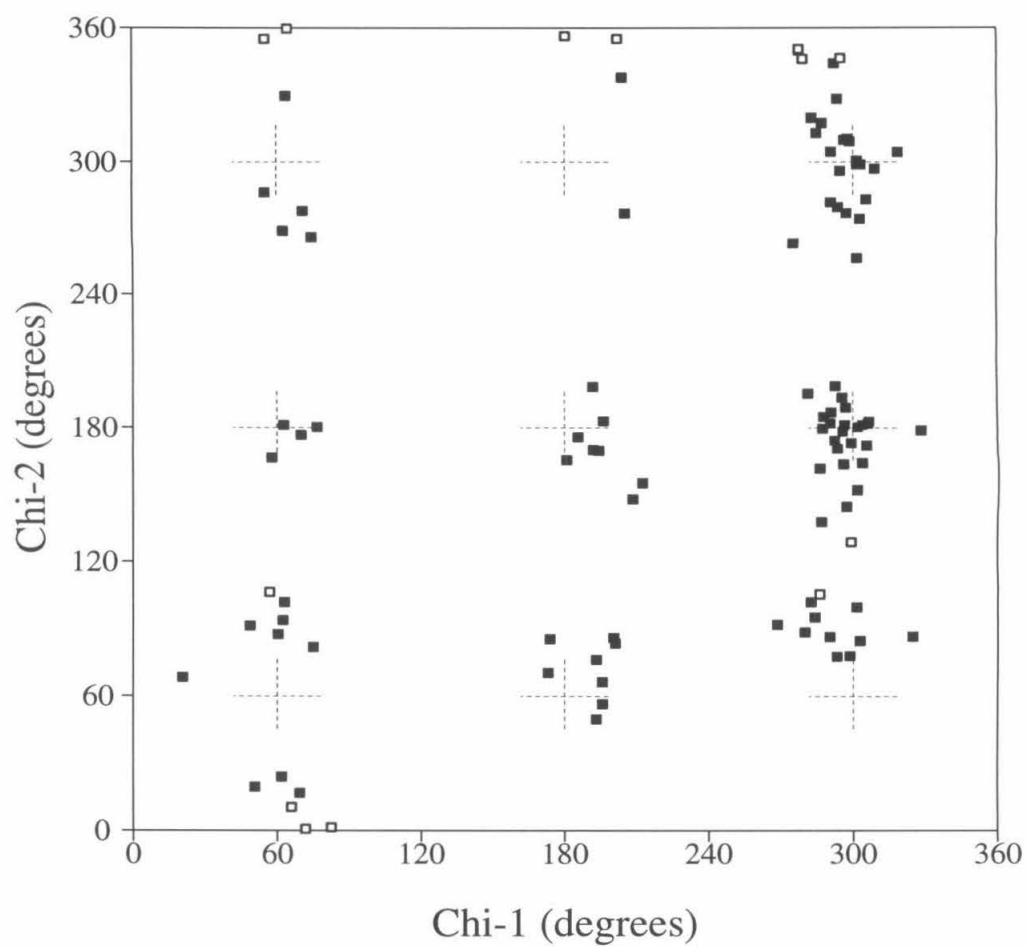


Plot statistics

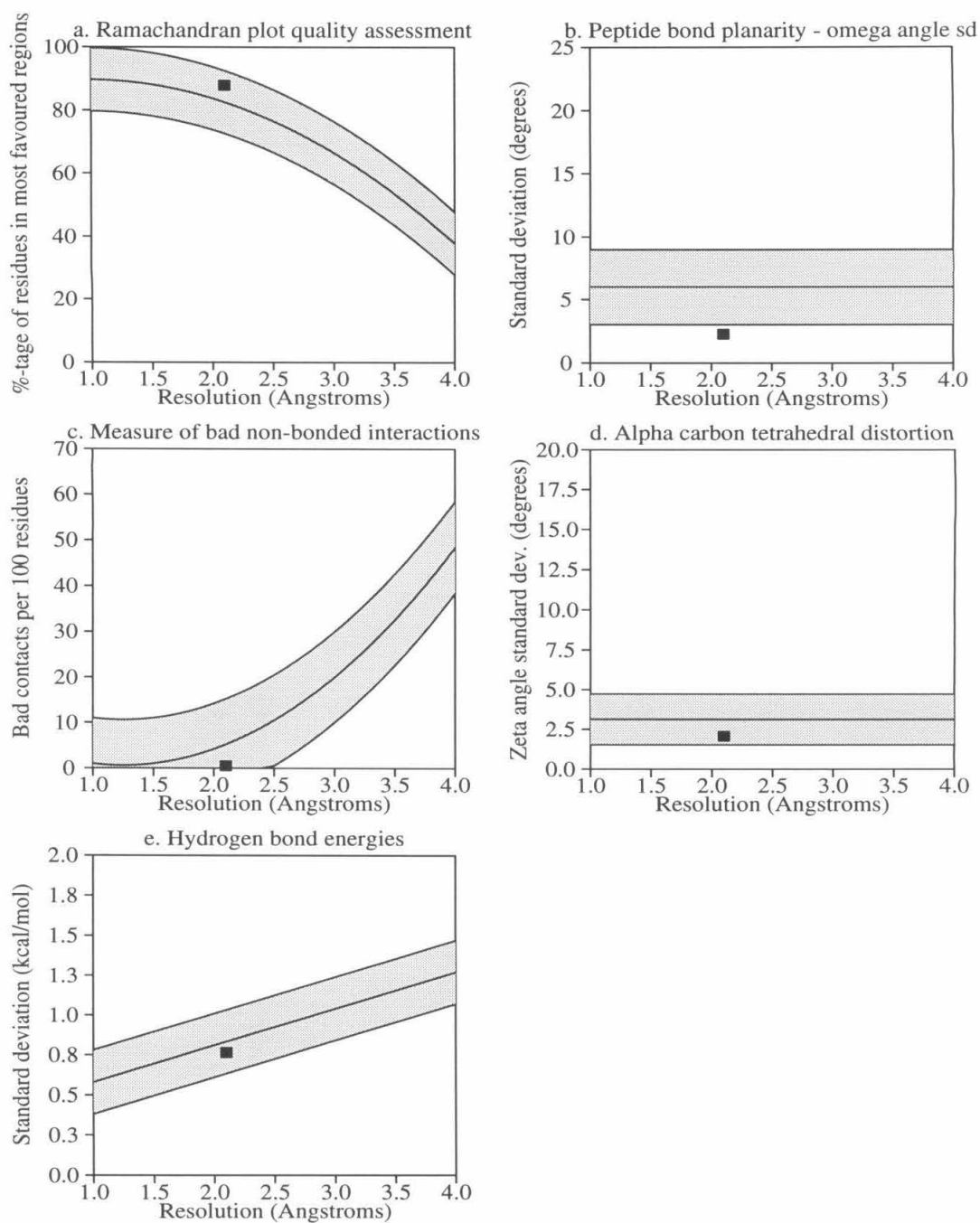
Residues in most favoured regions [A,B,L]	153	87.9%
Residues in additional allowed regions [a,b,l,p]	21	12.1%
Residues in generously allowed regions [\sim a, \sim b, \sim l, \sim p]	0	0.0%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	174	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	8	
Number of proline residues	16	

Total number of residues	202	

B.

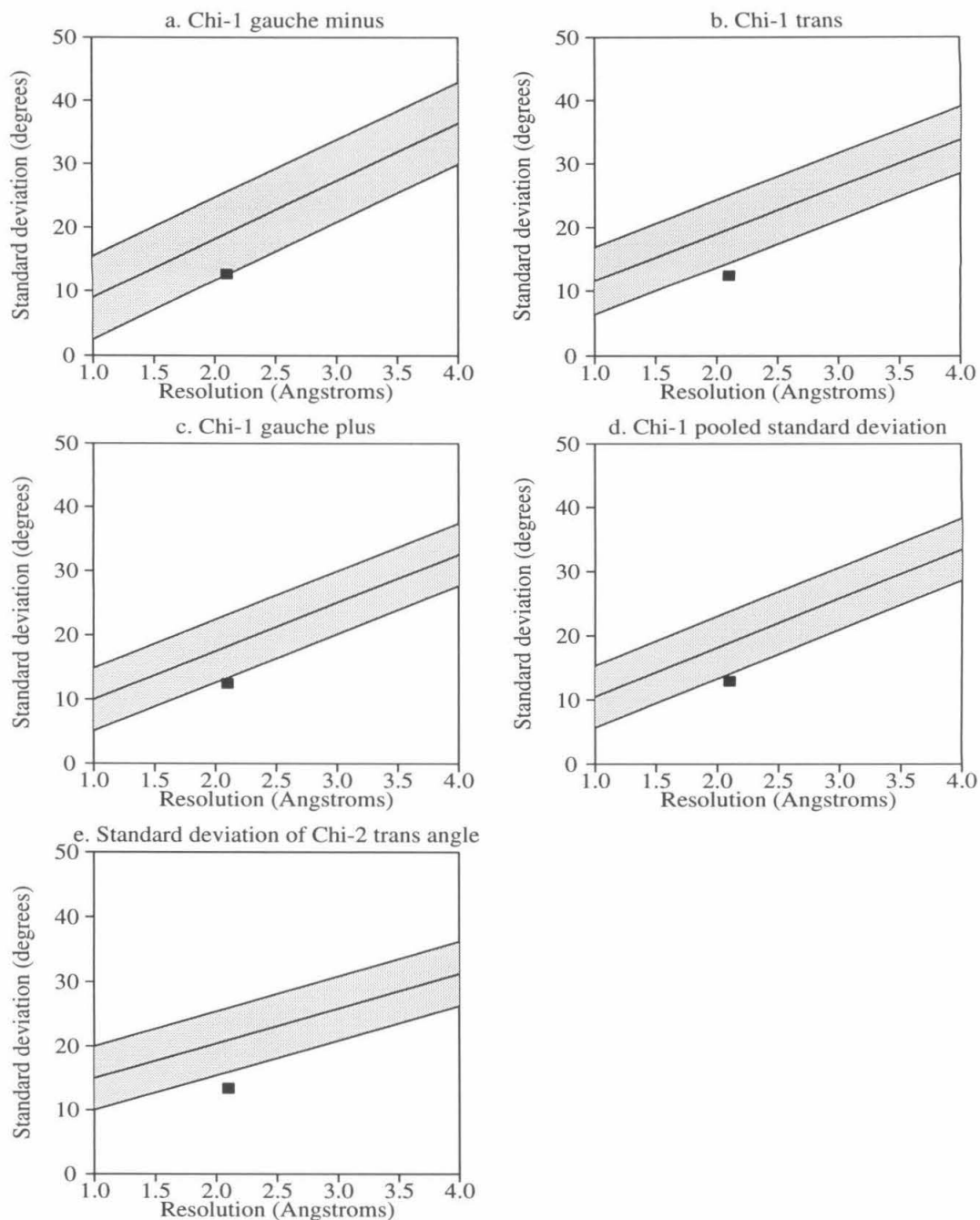
C.



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. %-tage residues in A, B, L	174	87.9	82.6	10.0	0.5
b. Omega angle st dev	200	2.3	6.0	3.0	-1.2
c. Bad contacts / 100 residues	1	0.5	5.2	10.0	-0.5
d. Zeta angle st dev	194	2.1	3.1	1.6	-0.7
e. H-bond energy st dev	101	0.8	0.8	0.2	-0.3

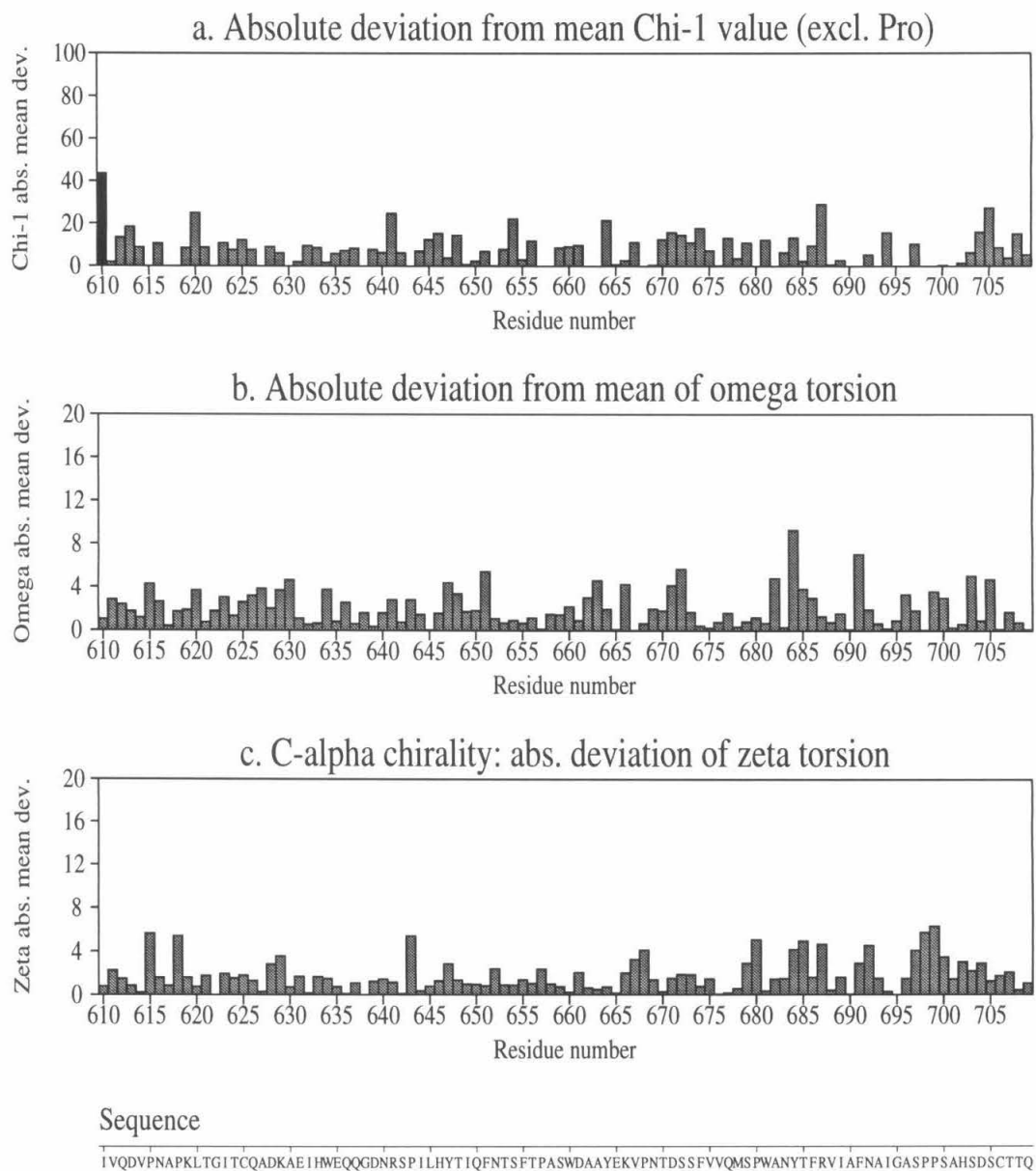
D.

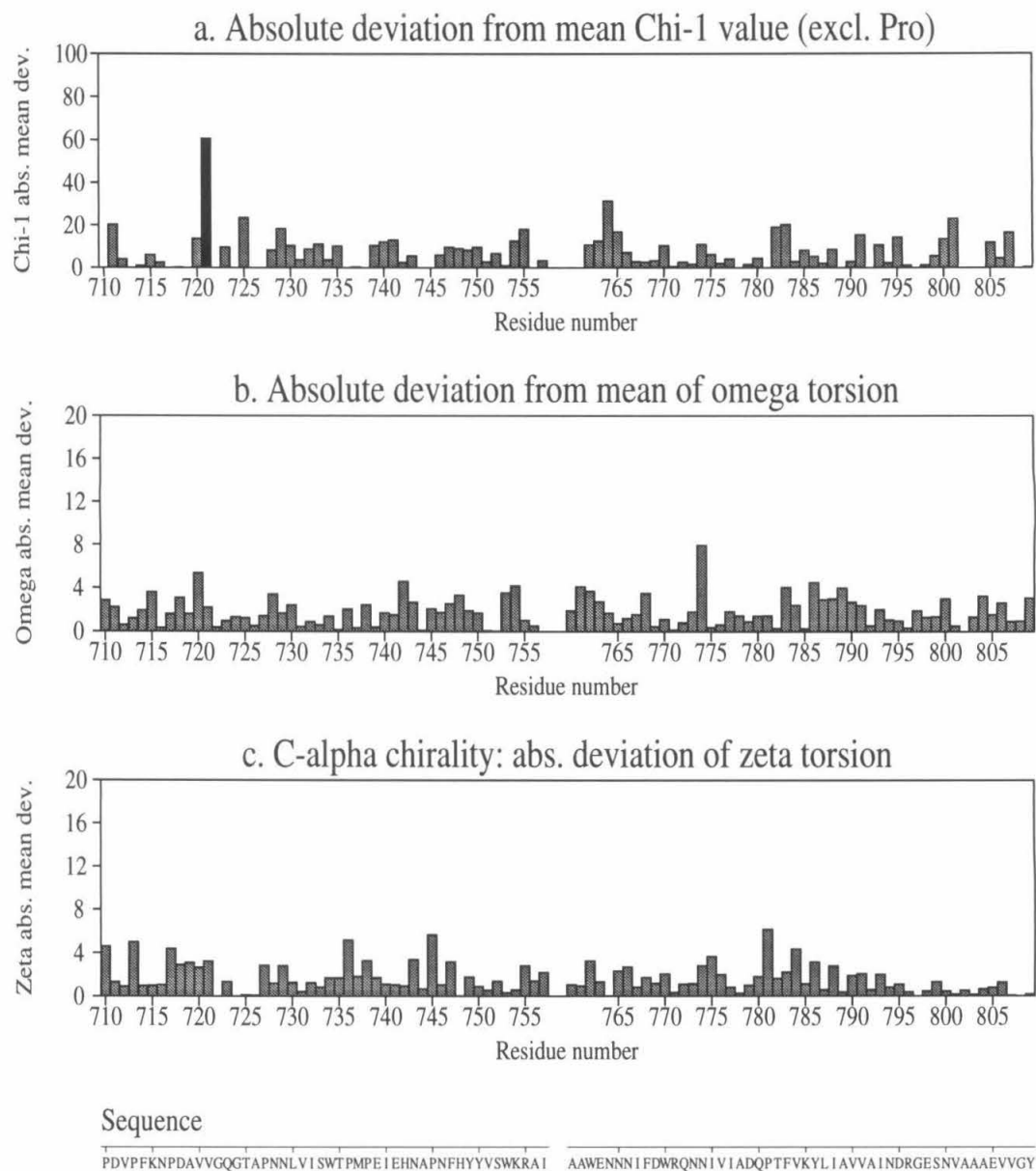


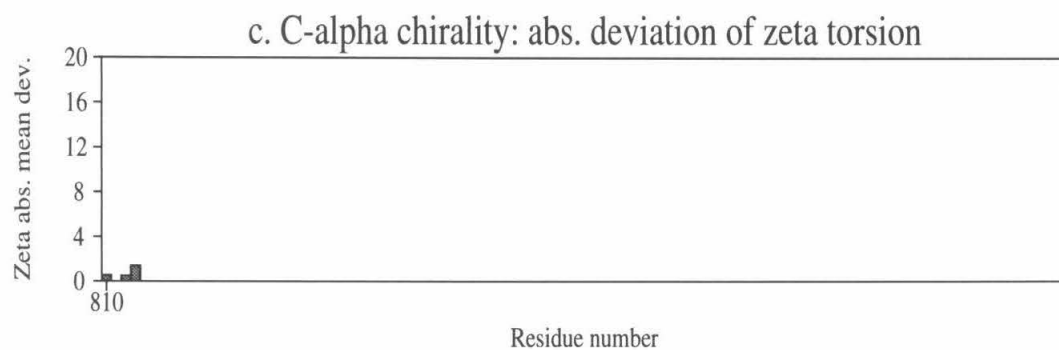
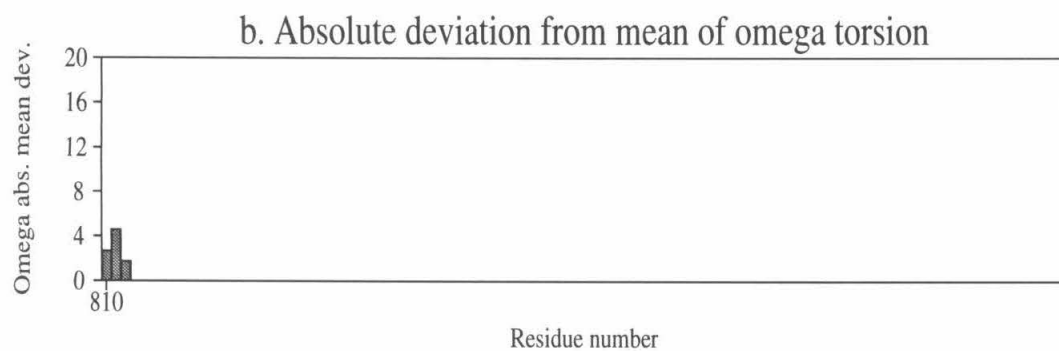
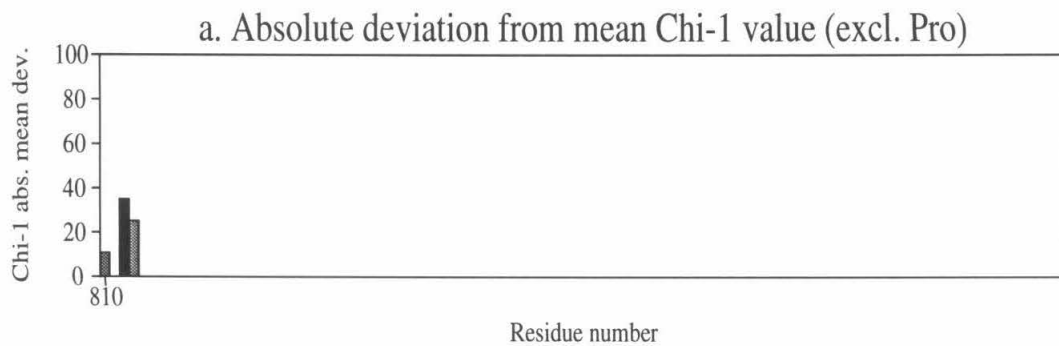
Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. Chi-1 gauche minus st dev	41	12.7	19.0	6.5	-1.0
b. Chi-1 trans st dev	36	12.5	19.7	5.3	-1.4
c. Chi-1 gauche plus st dev	78	12.5	18.3	4.9	-1.2
d. Chi-1 pooled st dev	155	13.0	18.9	4.8	-1.2
e. Chi-2 trans st dev	34	13.4	20.9	5.0	-1.5

E.



F.

G.

Sequence
SGED

Refinement cycle 5. The agreement between the refinement cycle 4 output model and the new difference maps was quite good. The rebuilding step produced a model containing all residues from Ile610 to Asp813, two N-acetylglucosamine (NAG) residues and 62 water molecules. Automated refinement was stopped after 15 cycles of conjugate gradient minimization because the R-factors were no longer being reduced. This refinement cycle produced a slight increase in both R_{cryst} and R_{free} (Table 3-1).

The refined model had tight geometry as measured by r.m.s. bond length and angle deviations and analysis with PROCHECK indicated that the stereochemistry was reasonable. Difference maps appeared to support the refined model and failed to point out any major modeling errors. This suggested that the only required modification was the placement of ordered water molecules. However, it seemed unlikely that additional water molecules would lower R_{cryst} from 26.6% to near 20%. It was apparent that this refinement process was not going to produce an acceptable model.

One factor that may have caused the refinement process to fail was the inclusion of data to 2.1 Å during the initial and subsequent refinement cycles. It was possible that the true positions of several model atoms were outside the radius of convergence permitted by the high resolution data. To test this possibility a “second” refinement process was started using an initial high resolution cutoff of 3.0 Å. In order to remove any model bias that might be present in the individual atomic B-factors, all individual temperature factors were set to 20.0 Å² and the overall B-value for the model was refined (Table 3-1, cycle 5). The final overall B for the model was 24.0 Å².

Refinement cycle 6. The refined model from cycle 5 was rebuilt. The new model contained all amino acids from Ile610 to Asp813 and two NAG residues. This model was

Table 3-2 Statistics for refinement cycles 6 to 21.

Refinement			Type of	Resolution		INPUT MODEL					OUTPUT MODEL				
Cycle	Refinement	(Å)	A. A.	Sugars	Waters	Other	R _{free}	R _{cryst}	r.m.s. Bond	r.m.s. Angle	R _{free}	R _{cryst}			
6	45CG	5.0 - 3.0	204	2	0	0	29.29	27.14	0.024	2.26	28.75	23.94			
	30CG	5.0 - 2.7					30.62	26.57	0.022	2.23	30.31	25.93			
7	45CG, 20AB	5.0 - 2.7	206	2	0	0	31.63	27.29	0.023	2.33	29.28	24.01			
8	60CG, 20AB	5.0 - 2.5	206	2	0	0	31.30	27.12	0.016	2.07	29.79	25.19			
9	60CG, 20AB	5.0 - 2.5	206	2	0	2-S	30.35	25.80	0.018	2.19	29.50	24.73			
10	60CG, 30AB	5.0 - 2.5	206	2	0	0	32.74	28.67	0.017	2.10	30.37	25.57			
11	30CG, 20AB	5.0 - 2.3	206	2	6	0	31.70	27.28	0.016	2.08	30.70	26.19			
12	30CG, 20AB	5.0 - 2.3	206	2	61	0	29.36	25.04	0.017	2.07	28.71	24.03			
13	60CG, 20AB	5.0 - 2.3	206	2	91	1-S	28.68	24.92	0.018	2.14	27.75	23.16			
14	60CG, 20AB	5.0 - 2.3	206	2	115	1-S	27.99	24.27	0.022	2.17	26.24	21.94			
15	60CG, 20AB	5.0 - 2.1	203	2	121	1-S	26.75	23.74	0.021	2.25	25.98	22.38			
16	60CG, 40AB	5.0 - 2.1	203	2	158	2-S	26.24	23.10	0.021	2.13	25.50	21.73			
17	60CG, 25UAB	5.0 - 2.1	203	2	169	2-S,1-M	26.30	23.19	0.020	2.27	24.72	20.45			
18	60CG, 25AB	5.0 -2.0	205	2	193	2-S,1-M	26.54	22.71	0.017	1.90	24.62	21.32			
19	120CG, 30AB	5.0 - 2.0	205	1	216	1-S,1N	25.05	22.37	0.015	1.79	23.96	21.05			
20	80CG, 30AB	5.0 - 2.0	205	2	254	1-S,1N	24.09	21.27	0.015	1.79	23.49	20.22			
21	130CG, 30AB	5.0 - 2.0	205	3	237	1-S,1N	23.72	20.43	0.015	1.90	23.48	20.17			

CG = cycles of Powell method conjugate gradient minimization; AB = cycles of minimization of restrained, individual, isotropic atomic temperature factors; UAB = cycles of minimization of unrestrained, individual, isotropic atomic temperature factors; 1B = cycles of minimization of an overall isotropic B factor for the structure; A.A. = amino acid residues; sugars = N-acetylglucosamine residues; waters = ordered water molecules; S = sulfate anion; M = magnesium cation; N = sodium cation; r.m.s. bond = root mean square deviation from ideal bond lengths; r.m.s. angle = root mean square deviation from ideal bond angles; R_{cryst} = $\Sigma_a(|F_o - F_c|) / (\Sigma_a F_o)$, where F_o and F_c are observed and calculated structure factor amplitudes and a_e (refinement-set reflections); R_{free} = $\Sigma_t (|F_o - F_c|) / (\Sigma_t F_o)$, where t ∈ (free-set reflections);

refined in two steps that used different cutoffs for the high resolution data (Table 3-2). The first step, consisting of 45 cycles of conjugate gradient minimization using data from 5-3.0 Å, produced very reasonable R-factors ($R_{\text{free}} = 28.75$, $R_{\text{cryst}} = 23.94$). This was followed by an additional 30 cycles of conjugate gradient minimization that included higher resolution data (5.0 - 2.7 Å). The second minimization step produced higher final R-factors ($R_{\text{free}} = 30.31$, $R_{\text{cryst}} = 25.93$), but this increase was at least in part due to the inclusion of the additional data.

Refinement cycle 7. In order to reduce model bias carried over from the failed first refinement process, the model was rebuilt using the initial SIRAS and solvent flattened SIRAS maps, as well as difference maps calculated with combined phases. The new model included residues Ile610 to Pro815 and two N-acetylglucosamines. Positional refinement and the refinement of individual restrained isotropic B-factors (automatic weighting) reduced the unbiased R-factor, R_{free} , a full percentage point, indicating that the information content of the model had been improved (Table 3-2).

Refinement cycle 8. Phases were extended to 2.5 Å because the final R_{cryst} ($R_{\text{cryst}} = 24.01\%$) in the previous cycle was quite reasonable. The difference maps used during this refinement cycle and all subsequent refinement cycles relied on the most recent refined model for all phasing information. The rebuilt model was minimized using a lower multiple (1.5x instead of 2x) of W_{ideal} , the value of W in equation 3-3 that results in an equal weighting of $E_{\text{EMPIRICAL}}$ and $E_{\text{X-RAY}}$. This step was taken in order to tighten up the model's geometry. The increase in final R_{free} and R_{cryst} values was probably due to the phase extension.

Refinement cycles 9 and 10. Neither of these cycles produced significant changes in the R-factors. In fact, both R-factors were a little worse after refinement cycle 10 than they had been at end of cycle 8. The models had good stereochemistry and most of the

[Fo-Fc] difference peaks were due to ordered water molecules. One peak was obviously not part of the protein and was too large and electron dense to be an ordered water molecule. This peak, which was within hydrogen bonding distance of His646, was modeled as a sulfate anion in cycle 9 because the crystallization conditions contained ammonium sulfate. A second sulfate was also placed during cycle 9 in another large difference peak. Both sulfates were removed during refinement cycle 10 because they were covered with negative [Fo-Fc] difference density and had severely distorted geometries.

Refinement cycle 11. No model errors were clearly evident in the difference maps. The stereochemistry of the model was analyzed with PROCHECK and several residues with energetically unfavorable side chain conformations were examined. Water molecules were placed in $+5\sigma$ [Fo-Fc] difference density if it was obvious the density could not belong to the protein and the position allowed the formation of chemically reasonable hydrogen bonds. These very stringent criteria resulted in the placement of six water molecules.

Refinement cycles 9 and 10 had failed to produce any improvement in R_{free} or R_{cryst} and relatively little rebuilding appeared to be needed during cycle 11. Further minimization using data from 5.0-2.5 Å was not likely produce a significant improvement in the model and so the high resolution data cutoff was extended to 2.3 Å (Table 3-2).

Refinement cycle 12. An additional 55 ordered water molecules with at least $+4\sigma$ [Fo-Fc] difference density and chemically reasonable hydrogen bonds were placed in the model. During this refinement cycle the value of the $E_{\text{X-RAY}}$ weight, W , started being set at twice W_{ideal} . The value of W was boosted in order to loosen up the geometry and to increase the importance of the crystallographic data in the minimization process. Automated refinement produced R-factors that were substantially lower than those of the

previous cycle ($R_{\text{cryst}} = 24.03\%$, $R_{\text{free}} = 28.71\%$) but the final value of R_{cryst} was still relatively high.

Refinement cycle 13. All waters with refined B-values greater than 50 \AA^2 were deleted during the rebuilding process. Thirty ordered water molecules were added to $+4\sigma$ [Fo-Fc] difference density peaks and the sulfate anion near His646 was placed in the model again because the orientation of the oxygen atoms was obvious. Model stereochemistry was examined with PROCHECK during the rebuilding process and a real space correlation coefficient was calculated for backbone and side chain atoms using the program O. The orientations of model peptide bonds were also examined using the Pep_flip utility in O. Both R-factors were significantly reduced during automated refinement (Table 3-2).

Refinement cycle 14. As in previous cycles, the rebuilding process relied heavily on [2Fo-Fc] and [Fo-Fc] difference maps and the analysis of model stereochemistry. There was still a poorly defined loop (Asp756-Ala761) in the structure that had weak [2Fo-Fc] difference density. Several of the residues in this loop had -4σ [Fo-Fc] difference density on or near residue atoms. The questionable amino acids in the loop were omitted and a simulated annealing omit map was calculated with data from $5.0\text{-}2.3 \text{ \AA}$. The loop residues were then placed in the model using the omit map. Calculating the omit map required the following steps: removal of the questionable region from structure, prestage, simulated annealing (initial temperature = 2,000), conjugate gradient minimization, and calculation of the [2Fo-Fc], ϕ_{calc} difference map. All water molecules with B-values greater than 50 \AA^2 were removed from the model and 24 new water molecules were placed using $+4\sigma$ [Fo-Fc] difference density peaks. Minimization produced substantial drops in the R-factors (R_{cryst} dropped by 1.22%, R_{free} dropped by 1.51%).

Refinement cycles 15 & 16. The rebuilding steps were carried out essentially as described for previous refinement cycles. Since the R-factors at the end of cycle 14 (R_{cryst}

= 21.94%, $R_{\text{free}} = 26.24\%$) were very reasonable, the high resolution data cutoff for automated refinement was extended to 2.1 Å during refinement cycle 15. The model used in refinement cycle 16 included amino acids Ile610 to Glu812, two NAG residues, 158 ordered water molecules and two sulfate anions. This model had good geometry and improved R-factors (Table 3-2).

Refinement cycle 17. Water molecules with a refined temperature factor greater than 55 Å² were deleted. The cutoff for an acceptable water molecule temperature factor was raised because water molecules deleted in previous cycles were reappearing in the new difference maps. A strong, positive [Fo-Fc] difference peak was located at the model's interdomain interface. This spherical difference density was within approximately 2.2 Å of five potential oxygen ligands that formed a site with approximate square pyramidal coordination geometry, suggesting that the difference density belonged to a metal ion. The ion was initially modeled as a magnesium cation with a charge of 2+ and estimated Lennard-Jones parameters. Energy minimization involved positional refinement and the unrestrained (W_B set to zero, equation 3-4) refinement of individual isotropic temperature factors.

Refinement cycle 18. The model was carefully examined during rebuilding for unusually high temperature factors that would be indicative of incorrectly placed atoms. The rebuilt model included residues Ile610 to Arg814, two N-acetylglucosamines, 193 ordered water molecules, two sulfate anions, and a Mg²⁺ cation. Starting with this cycle, individual reflections were weighted during the minimization process using the equation for w_h (from equation 3-3) given below:

$$w_h = 1.0 / (10.0 - 85.0 \times (S()/2.0 - 0.16667))^2 \quad (3-6)$$

where $S()$ represents the length of the reciprocal lattice vector for each reflection. This weighting scheme was designed to increase the weight of high resolution reflections and

decrease the weight of low resolution reflections such that data at 5.0 Å and 2.1 Å make approximately equal contributions to the E_{X-RAY} term. The high resolution data cutoff was increased to 2.0 Å for automated refinement, which included positional refinement and the refinement of restrained individual temperature factors. During the temperature factor refinement, the weight (W_B) of the restraining term E_R (equation 3-4) was set to twice the value automatically determined by X-PLOR in order to better meet the target deviations in atomic B-values. W_B was similarly inflated in all subsequent refinement cycles. After the temperature factor refinement, the B-value of the metal ion was set to the average of the oxygen ligand B-values and the occupancy of the ion was refined to the following values: $Mg^{2+} = 75.7\%$, $Ca^{2+} = 51.8\%$, $Na^+ = 78.1\%$.

Refinement cycle 19. Several corrections were made during this rebuilding step: all water molecules with $B > 55 \text{ Å}^2$ were deleted; 23 new water molecules were manually placed; one sulfate was removed because it had high B-values and poor geometry; the Mg^{2+} was changed to Na^+ ; and the second NAG was removed because it had high B-values and strong negative [Fo-Fc] difference density. In order to obtain metal-ligand distances not biased by electrostatic and van der Waals forces, the charge on the Na^+ ion was set to zero and the ion's Lennard-Jones parameters were set to values near zero. Minimization produced a significant drop in R_{free} indicating that the information content of the model had been improved. The B-value of the ion in the refined model was set to the average B-value of its oxygen ligands and the occupancy of the Na^+ ion was refined to 77%.

Refinement cycle 20. All water molecules with B-values greater than 56 Å^2 were deleted from the initial model and 38 new waters were manually placed in $+4\sigma$ [Fo-Fc] difference peaks. The N-acetylglucosamine residue at the second utilized N-linked glycosylation site (Asn683) was finally placed. Strong positive density attached to the Asn683 side chain had been observed previously but this was the first [Fo-Fc], ϕ_{calc} map in

which the orientation of the sugar ring was clear. Positional refinement and restrained isotropic temperature factor refinement produced a significant drop in R_{free} (Table 3-2).

Refinement cycle 21. During the last few refinement cycles, water molecules deleted during one cycle for having too large a B-value were frequently reappearing in the following cycles. These molecules had $+4\sigma$ [Fo-Fc] difference density and reasonable hydrogen bonding partners. In this refinement cycle, the cutoff for a water molecule atomic B-value was raised to 61 \AA^2 and new water molecules were placed in $+4\sigma$ [Fo-Fc] difference density. Other changes made during this rebuilding step included: fixing the orientation of Asn and Gln side chain amide groups and placing the second NAG residue at the Asn652 N-linked glycosylation site. Positional and restrained atomic B-factor refinement failed to produce any further improvement in the model as judged by the R-factors (Table 3-2).

The Final Model

The final refined model includes amino acid residues Ile610 to Arg814, three N-acetylglucosamine residues, 237 water molecules, a sulfate anion, and a Na^+ cation. The reliability of this model was examined using several criteria.

Common indicators of overall reliability. The final model has good geometry (r.m.s. bond and angle deviations of 0.015 \AA and 1.9° respectively) and reasonable R-factors ($R_{\text{cryst}} = 20.17\%$ $R_{\text{free}} = 23.48\%$) for all data from 5 to 2.0 \AA . The unbiased R-factors (R_{free}) for recently published structures typically have values which range from the low thirties to the middle twenties (expressed as a percentage) and thus the R_{free} value for this structure is particularly low.

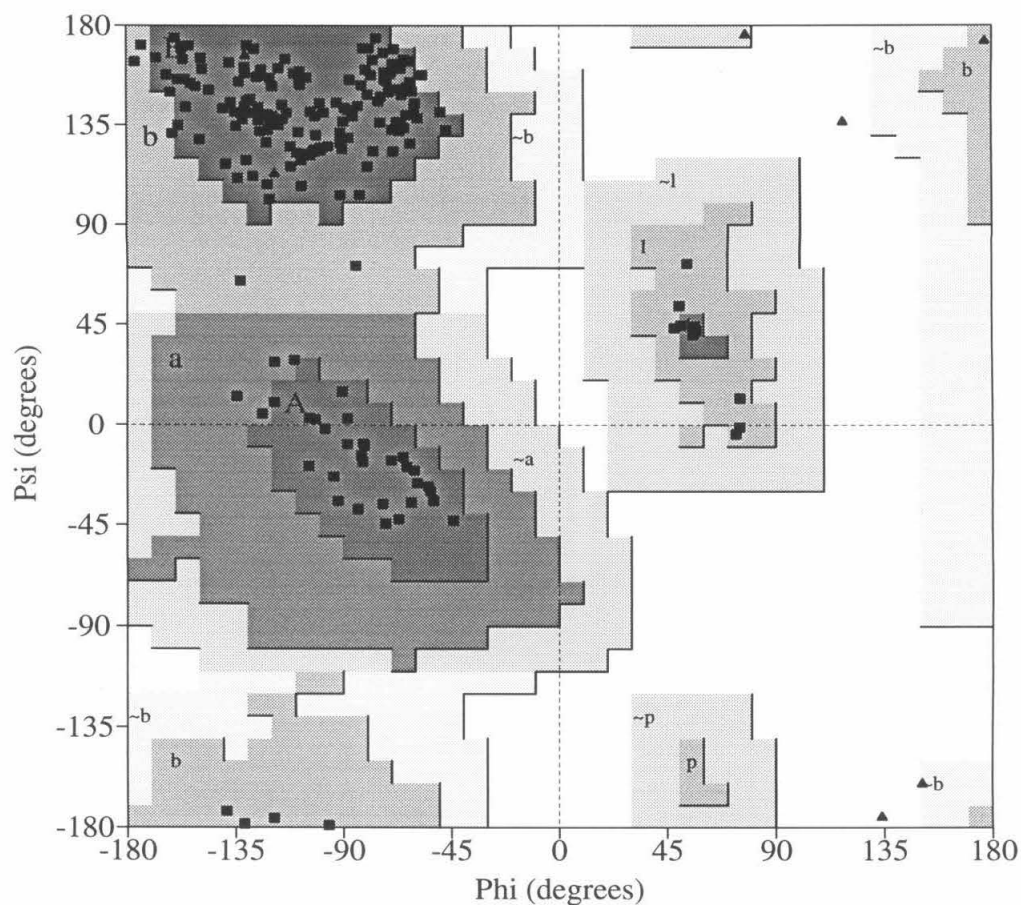
However, the absolute value of R_{free} is not currently used to judge the reliability of a model because it is not yet clear what values are reasonable. Some structures with good

conventional R -factors (R_{cryst}) have relatively high values for R_{free} . While the magnitude of the difference between R_{cryst} and R_{free} is, in some part, a reflection of model quality, it is also influenced by noise in the data, the completeness of the atomic model, and the observable to parameter ratio¹³. None of these factors were considered to be a problem with the NgFn1,2 structure.

Geometry. Analysis of the model's geometry confirmed that main chain and side chain stereochemistry is reasonable. The phi, psi angles for all non-glycine residues are within allowed regions with over 88% of these residues falling within most favored regions (Figure 3-2A). PROCHECK uses several criteria to evaluate the stereochemistry of the backbone. These are shown in Figure 3-2C and include Ramachandran plot quality, peptide bond planarity, bad non-bonded interactions, carbon- α tetrahedral distortion, and hydrogen bond energies. The values for all of these criteria were better than average (determined from a database of well refined structures) for the NgFn1,2 structure, with peptide-bond planarity being particularly good.

Figure 3-2 (following pages) Modified Procheck output from analysis of final model. (A) Ramachandran plot of model phi, psi angles. An analysis of 118 structures solved to at least 2.0 Å resolution with an R_{cryst} of no greater than 20% suggests that a good quality model will have over 90% of its phi, psi angles in the most favored regions. The most favorable regions of phi, psi space have the darkest shading in the diagram. Less favorable regions have less shading. (B) Plot of chi-1 versus chi-2. The dashed crosses represent the ideal gauche and trans regions for the chi-1 and chi-2 dihedral angles. The crosses extend to one standard deviation from the ideal values and points that are more than 2.5 standard deviations from the ideal are shown as empty squares. This plot includes 117 of the 205 residues present in the model. Twelve of these residues are represented as empty squares. This program artificially inflates the number of residues with poor chi-1, chi-2 angles by including asparagine residues in the analysis. These residues do not have ideal chi-2 values at 60, 180 and 300°. (C) Assessment of quality of polypeptide backbone parameters. These plots present the overall quality of the polypeptide backbone's stereochemistry in terms of the following parameters: phi and psi dihedral angles, peptide bond planarity, bad non-bonded interactions, carbon- α tetrahedral distortion, and hydrogen bond energies. The line through the middle of the shaded area in each plot is the mean value expected for a well refined structure. The shaded area itself represents one standard deviation about the mean. The value for the NgFn1,2 structure is represented by a shaded square. (D) Assessment of the quality of side chain parameters. These plots present the overall reliability of the placement of side chains in the model in terms of chi-1 and chi-2 angles. The shaded band in each plot represents one standard deviation about the mean value (central line) expected for a well refined model. The value for the NgFn1,2 model is represented by a shaded square. (E, F and G) Plot of stereochemical parameters on a per residue basis. Deviations in chi-1 angles, peptide bond planarity and carbon- α chirality are plotted on a per residue basis allowing the identification of problematic areas in the model. Highlighted residues are those that deviate by more than 2.0 standard deviations from ideal.

A.

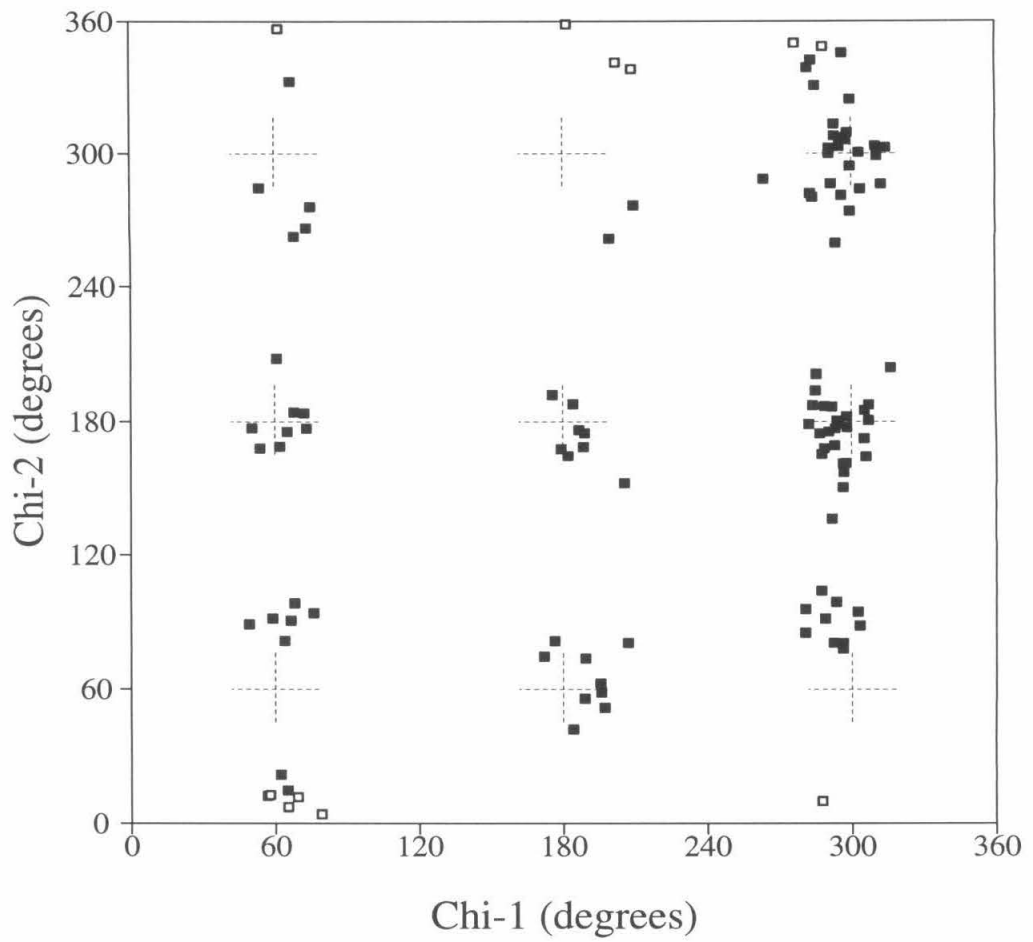


Plot statistics

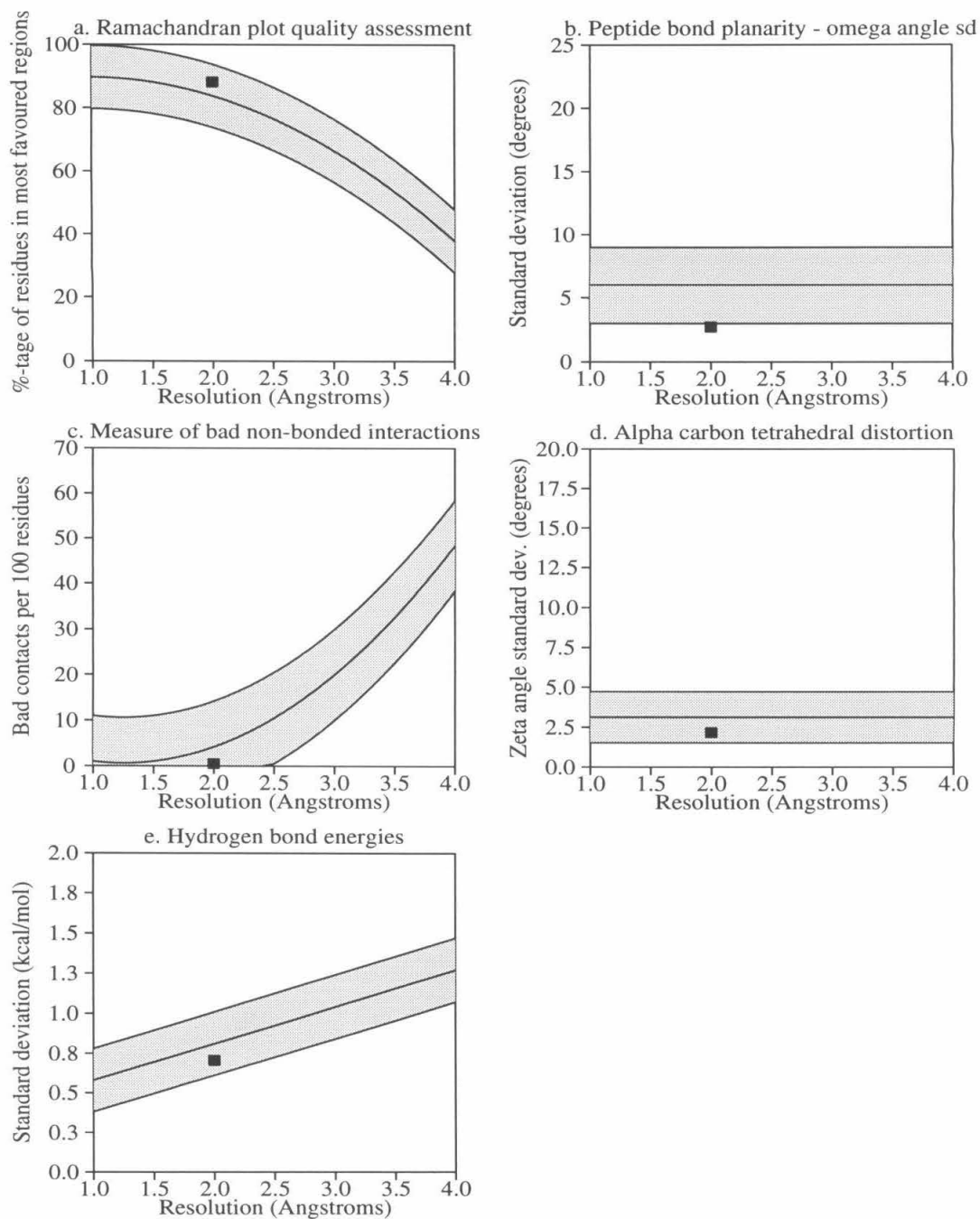
Residues in most favoured regions [A,B,L]	157	88.2%
Residues in additional allowed regions [a,b,l,p]	21	11.8%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	178	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	8	
Number of proline residues	17	

Total number of residues	205	

B.

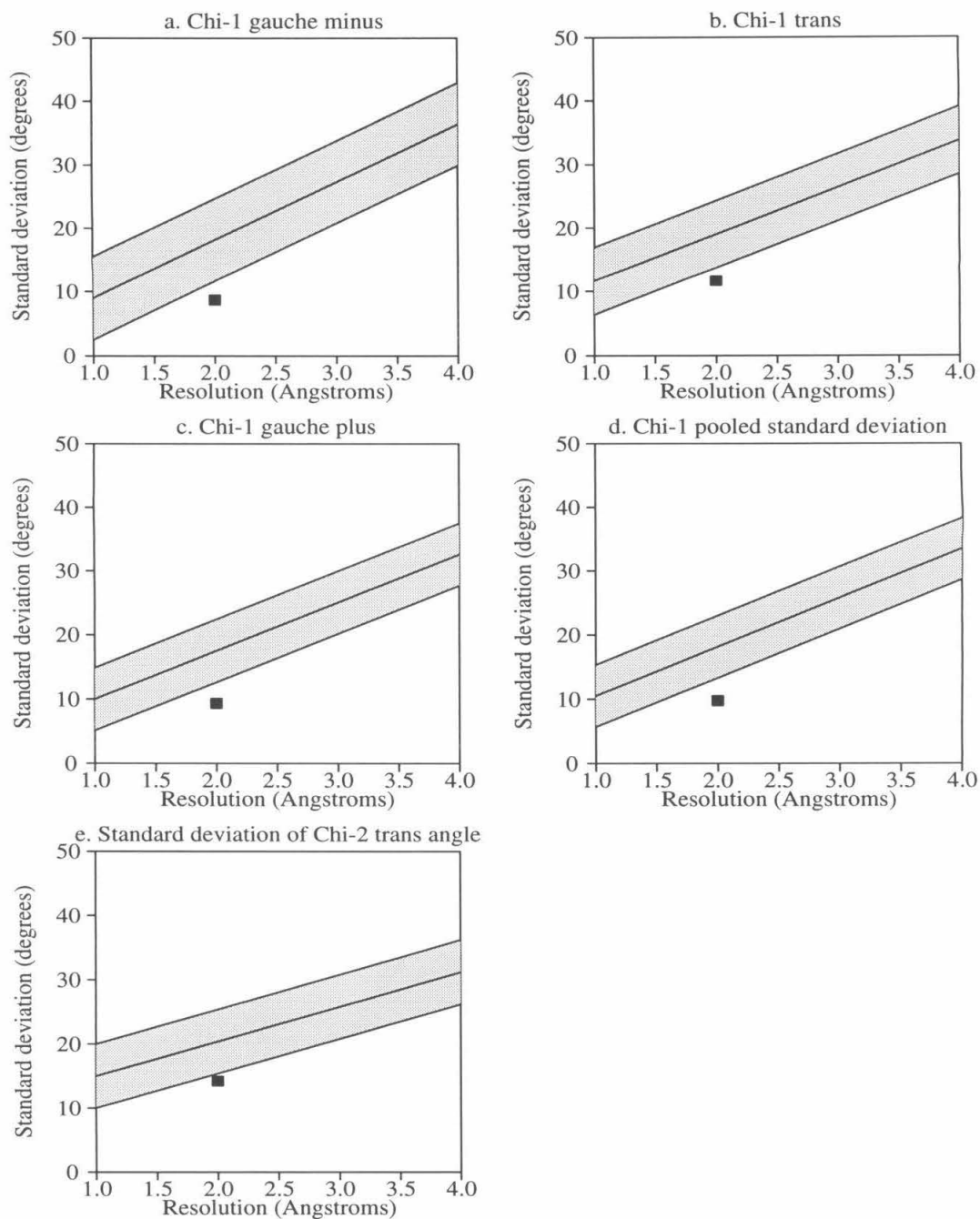
C.



Plot statistics

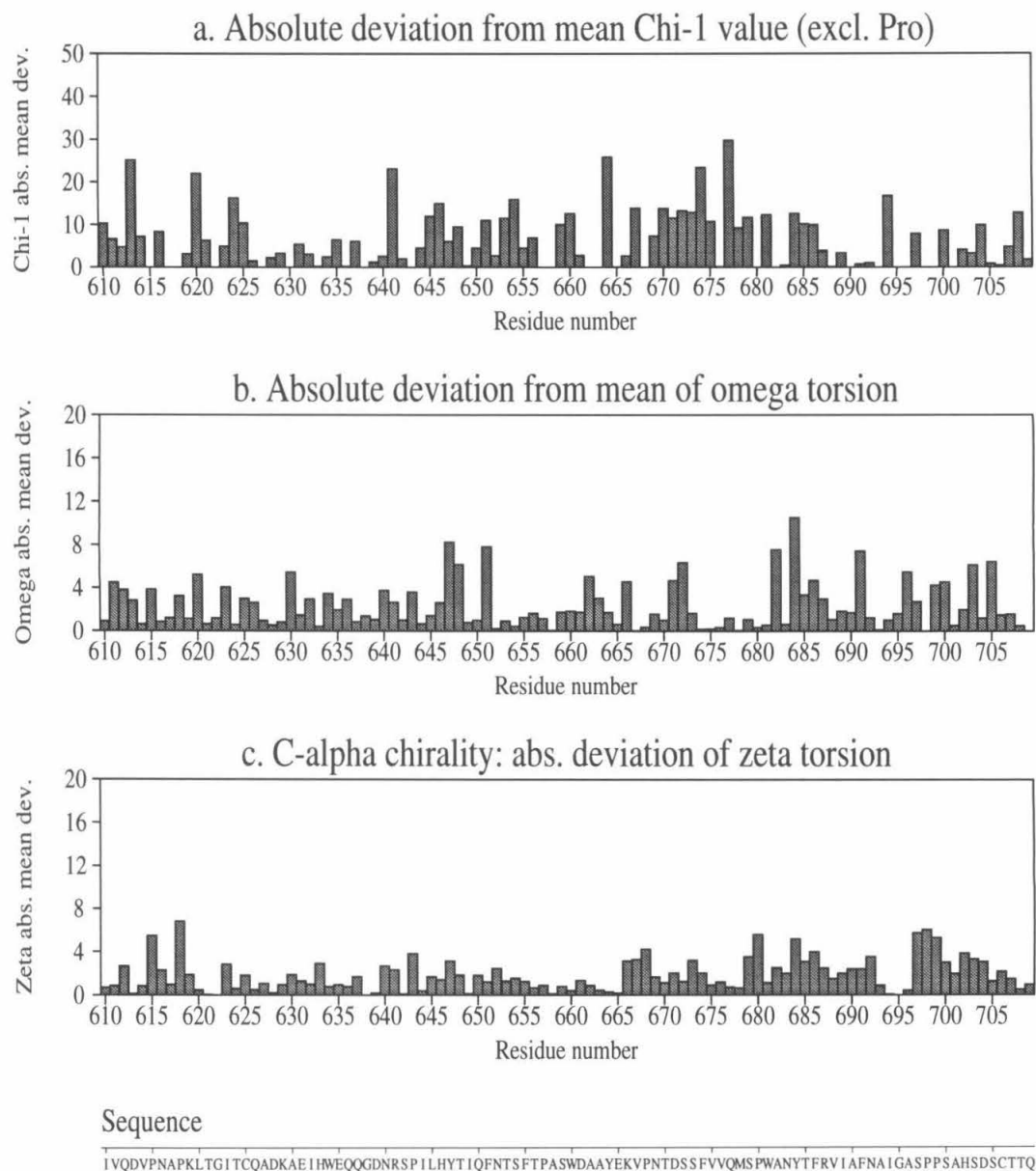
Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. %-tage residues in A, B, L	178	88.2	83.8	10.0	0.4
b. Omega angle st dev	204	2.7	6.0	3.0	-1.1
c. Bad contacts / 100 residues	1	0.5	4.2	10.0	-0.4
d. Zeta angle st dev	197	2.2	3.1	1.6	-0.6
e. H-bond energy st dev	100	0.7	0.8	0.2	-0.5

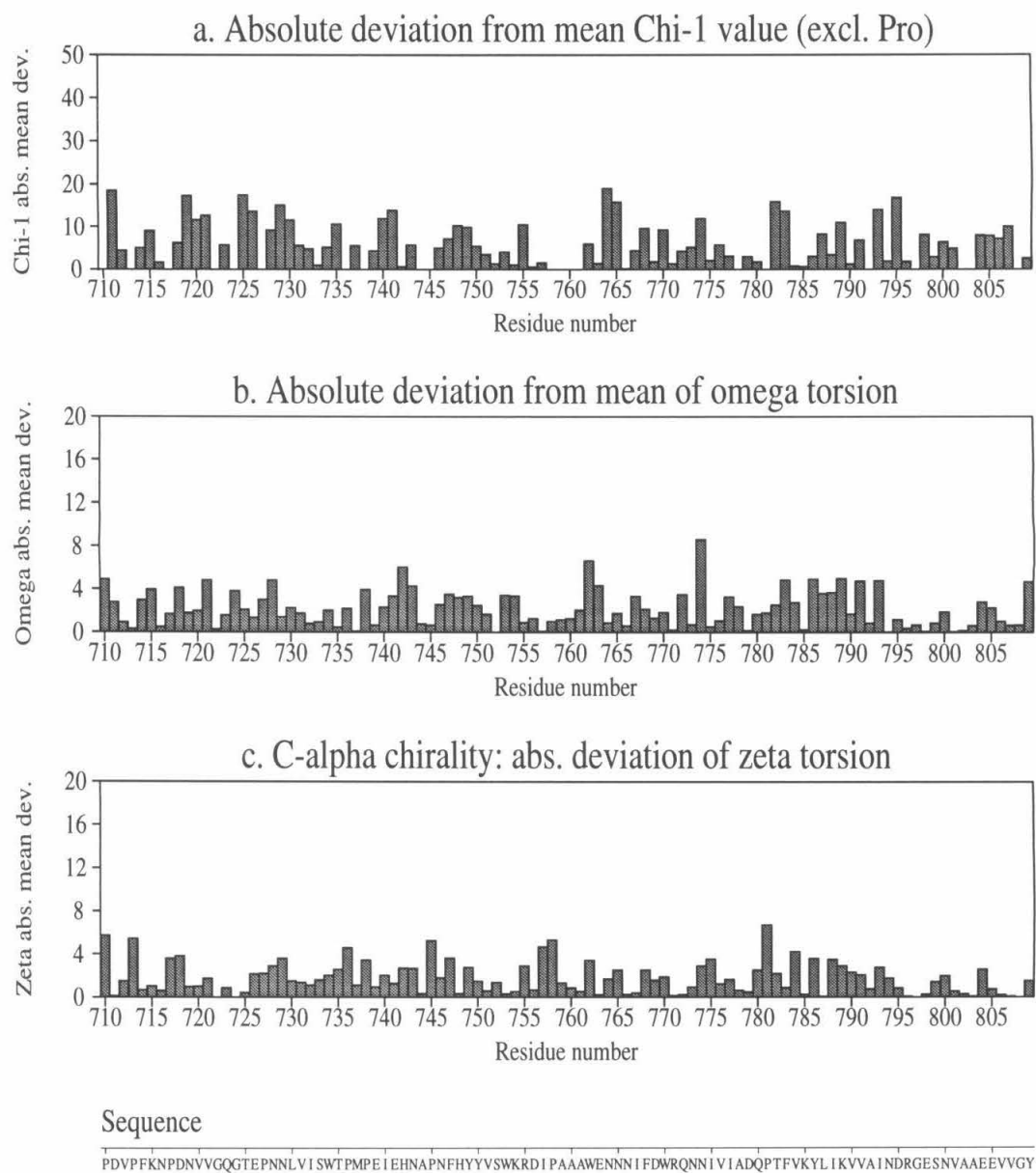
D.

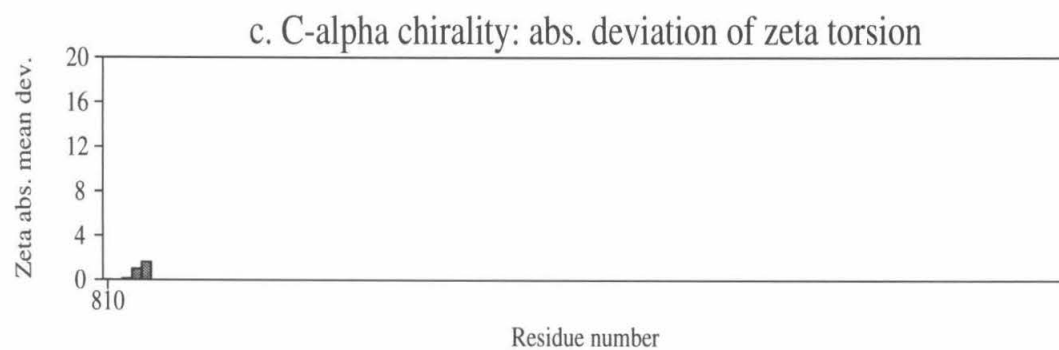
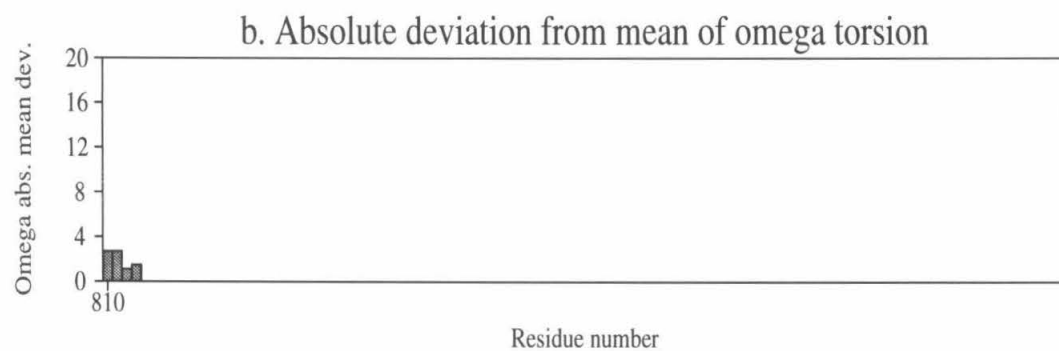
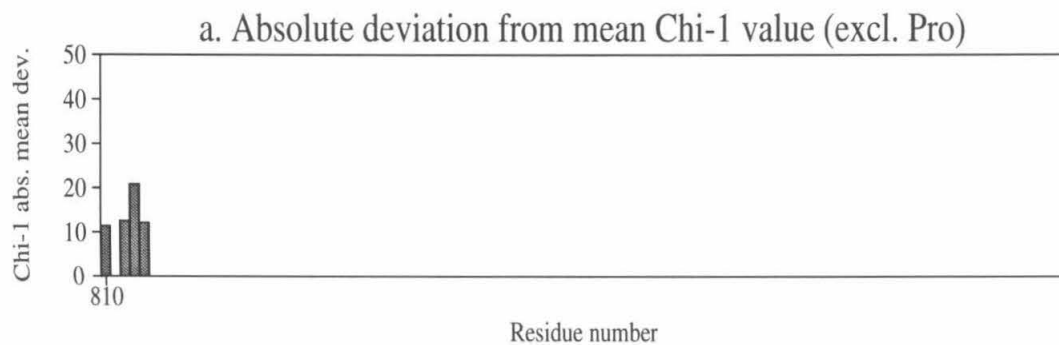


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. Chi-1 gauche minus st dev	43	8.7	18.1	6.5	-1.5
b. Chi-1 trans st dev	38	11.6	19.0	5.3	-1.4
c. Chi-1 gauche plus st dev	80	9.3	17.5	4.9	-1.7
d. Chi-1 pooled st dev	161	9.8	18.2	4.8	-1.7
e. Chi-2 trans st dev	42	14.2	20.4	5.0	-1.2

E.

F.

G.

Sequence
SGEDR

The stereochemistry of model side chains is checked by analyzing the side chain dihedral angles χ_1 and χ_2 . Due to stereochemical constraints, 60° , 300° (gauche) and 180° (trans) dihedral angles are most energetically favorable. PROCHECK produces a plot of χ_1 vs. χ_2 which is somewhat analogous to a Ramachandran plot. PROCHECK also presents an overall analysis of χ_1 and χ_2 values for the structure (Figure 3-2D) which shows that NgFn1,2 has better than average χ_1 and χ_2 values (compared to database of well refined structures).

A combination of backbone and side chain parameters are evaluated and used to rate the stereochemistry of each amino acid (Figure 3-2E-G). These parameters include χ_1 values, peptide bond planarity and carbon- α tetrahedral distortion.

Peptide bond orientations were monitored with the Pep_flip utility in the program O. For each residue i , Pep_flip searches a database of well refined structures for pentapeptides that have the same backbone conformation as the model fragment containing residues $i-2$ to $i+2$. The r.m.s. deviation between model and database backbone oxygen coordinates for residue i is used as an index of fit. Pep_flip values of greater than 2.5 \AA are considered suspect, with r.m.s. deviations greater than 3 \AA indicating that model and database oxygens are pointed in opposite directions. The Pep_flip analysis of the final model is shown in Figure 3-3A,B. There are four backbone carbonyls in the NgFn1,2 structure with Pep_flip values greater than 3.0. Two of these oxygen atoms were involved in hydrogen bond formation suggesting that they were correctly placed. None of the four oxygens had unusually high B-values or associated positive or negative [Fo-Fc] difference density that would indicate they were incorrectly placed.

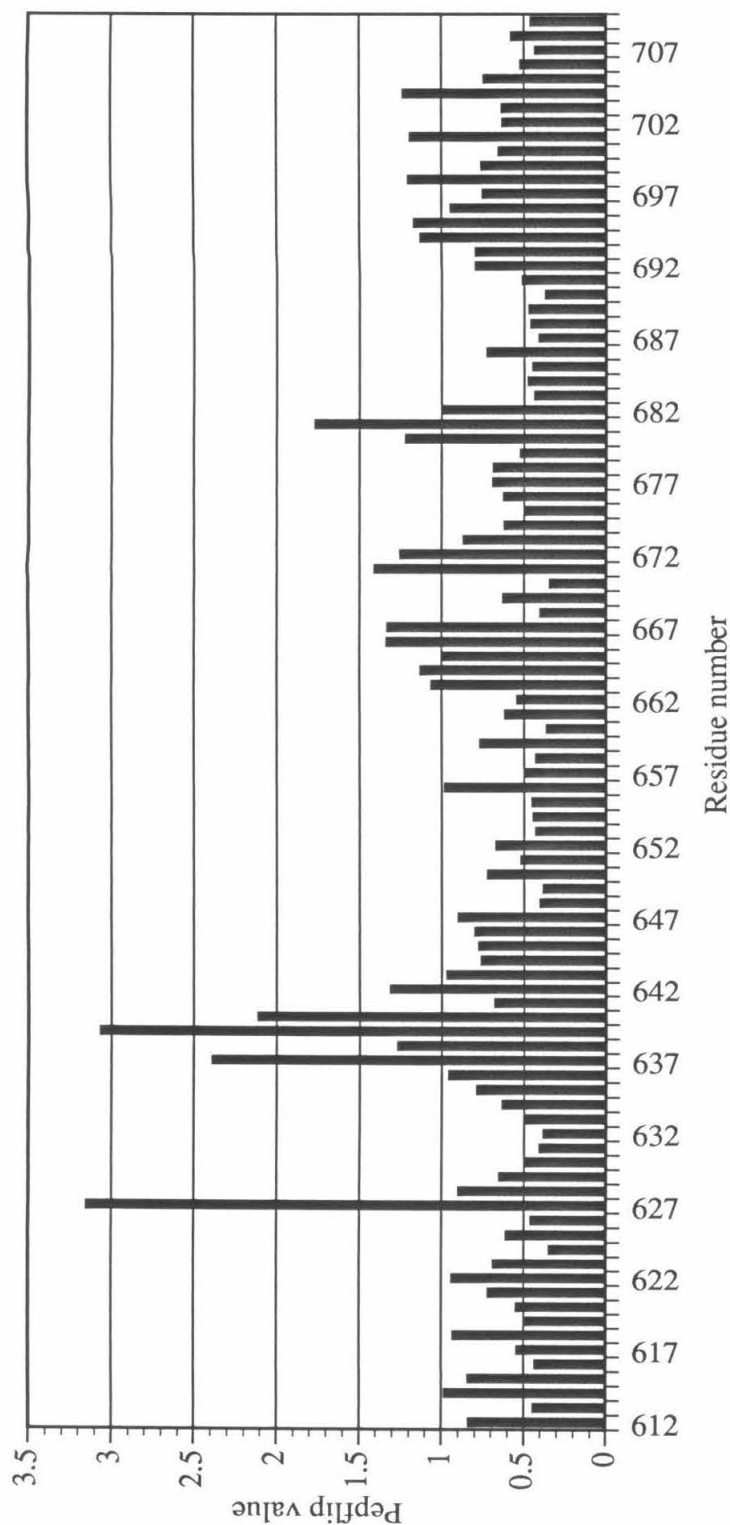


Figure 3-3A Plot of Pep_flip value versus residue number. The program O was used to generate a Pep_flip value for each residue. The Pep_flip value is actually the r.m.s. deviation between model backbone oxygen coordinates and those present in a database of well refined structures. For each residue i , Pep_flip searches the database for pentapeptides that have the same backbone conformation as the model fragment containing residues $i-2$ to $i+2$. Pep_flip values are not calculated for the two residues at the amino and carboxy termini of the structure. Values of greater than 2.5 \AA are considered suspect with r.m.s. deviations greater than 3 \AA indicating that model and database oxygens are pointed in opposite directions. Residues from NgFn1 (Gln612 to Gln709) are shown.

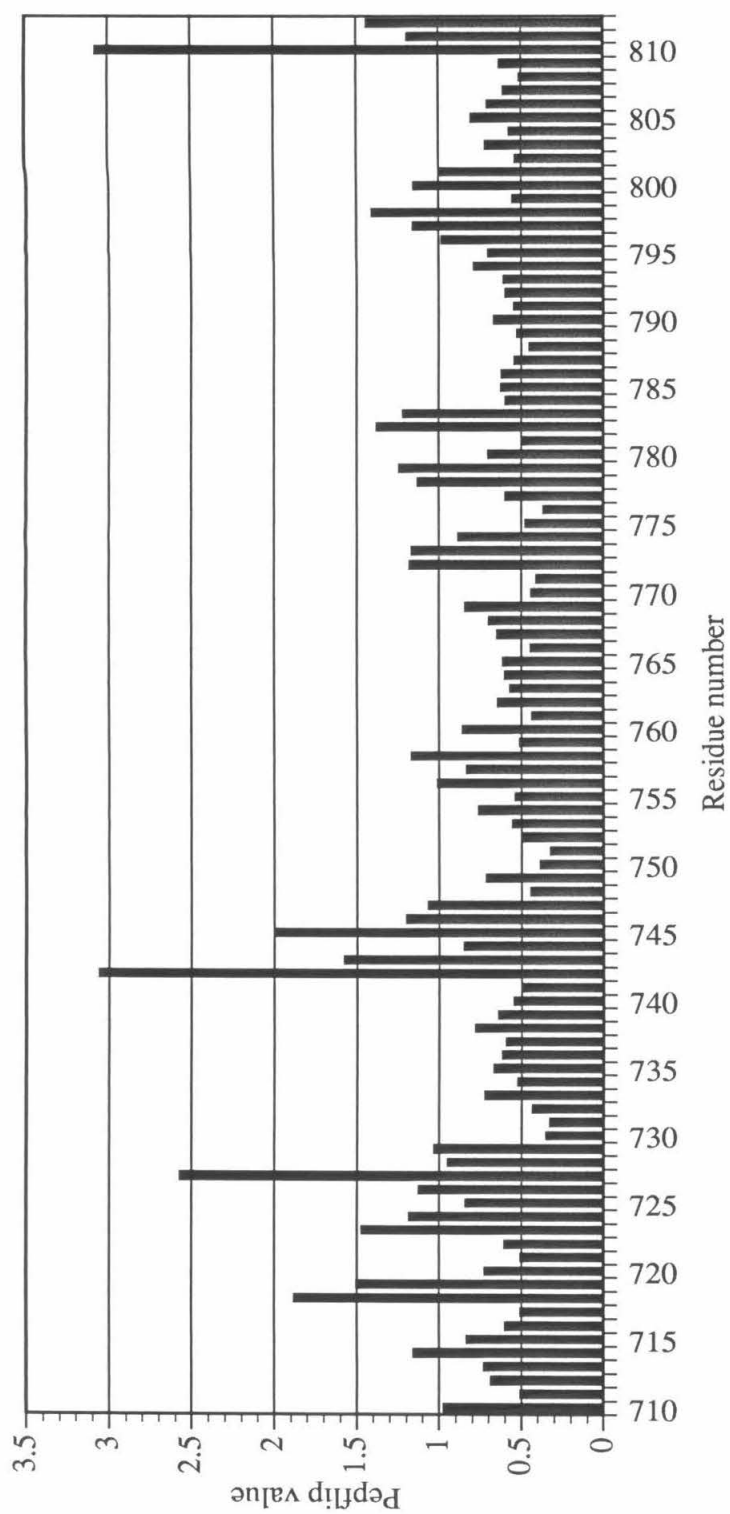


Figure 3-3B Plot of Pep_flip value versus residue number. Residues from NgFn2 (Pro710 to Arg814) are shown.

Real space correlation coefficient. Another method of evaluating the structure on a per residue basis involves measuring how well each residue fits the observed electron density¹⁴. This real space fit can be calculated with the Rs_fit utility in the program O, which expresses the quality of fit as a correlation coefficient. Figures 3-4 and 3-5 show the real space fit of the final model to three electron density maps: a SIRAS map calculated to 2.6 Å, a 2.6 Å solvent flattened SIRAS map, and the final [2Fo-Fc], ϕ_{calc} difference map. The real space fit was calculated separately for backbone atoms (Figure 3-4) and side chain atoms (Figure 3-5).

Analysis of temperature factors. Scattering power, which measures how effectively an atom scatters x-rays, is a function of electron density. An atom's scattering power is reduced by thermal motion because the electron cloud of the atom is spread over the volume covered by the motion. This reduces the observed electron density for that atom. Atomic temperature factors, which are used to model the effects of thermal motion, are large ($>60 \text{ Å}^2$) for atoms placed in weak electron density and low ($<25 \text{ Å}^2$) for atoms placed in strong density. Thus, an analysis of temperature factors can be used to evaluate structural mobility with anomalously high temperature factors frequently indicating errors in the model.

The mean atomic temperature factor for all atoms in the NgFn1,2 model is 29.4 Å^2 with backbone and side chain atoms having average B-values of 28.2 Å^2 (0.92 Å^2 r.m.s. deviation) and 30.6 Å^2 (1.84 Å^2 r.m.s. deviation), respectively. It was somewhat surprising that this structure had such a high average B-value since it diffracted to high resolution and the data were collected from frozen crystals. However, this average temperature factor is consistent with the fact that the high resolution data (data past 2.7 Å) were quite weak and could only be collected at a synchrotron.

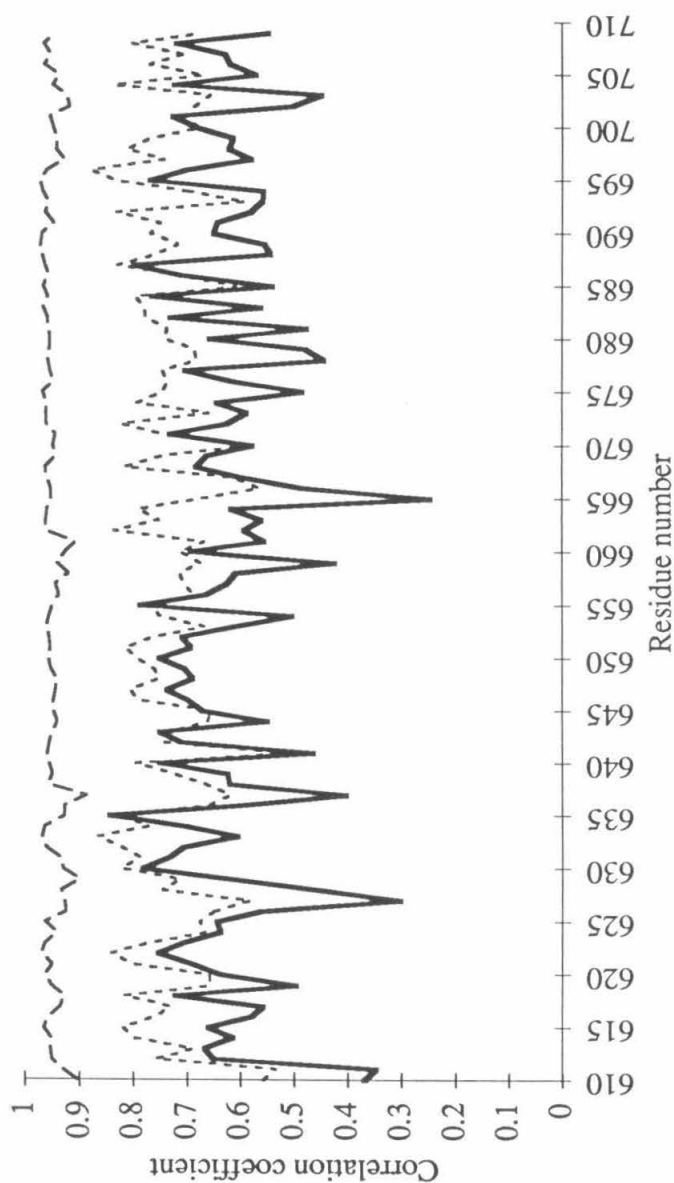


Figure 3-4A Plot of backbone real space correlation coefficient as a function of residue number. The real space fit of the final model was calculated with the Rs_fit utility in the program O, which expresses the quality of fit as a correlation coefficient. The real space fit was calculated separately for three electron density maps: a SIRAS map calculated to 2.6 Å (thick solid line), a 2.6 Å solvent flattened SIRAS (thin with short dashes), and the final $[2F_o - F_c]$, ϕ_{calc} map (thin with long dashes). Residues Ile610 to Gln709 are shown.

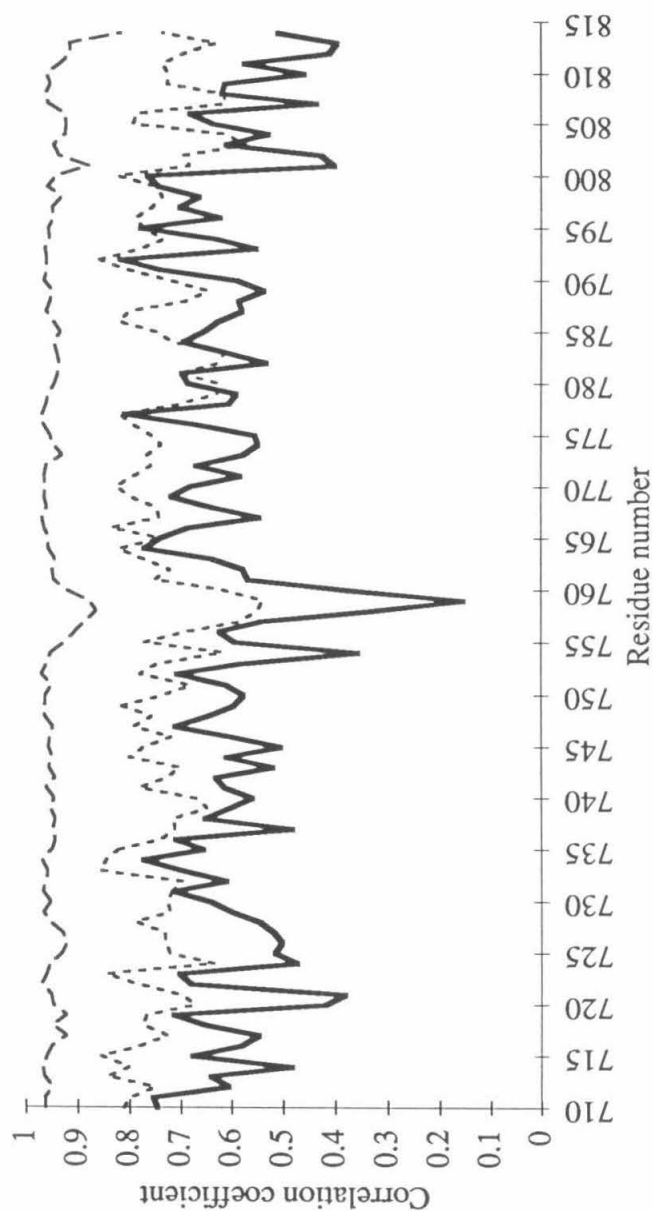


Figure 3-4B Plot of backbone real space correlation coefficient as a function of residue number. The real space correlation coefficient was calculated separately for three electron density maps: a SIRAS map calculated to 2.6 Å (thick solid line), a 2.6 Å solvent flattened SIRAS (thin with short dashes), and the final [2Fo-Fc], ϕ_{calc} map (thin with long dashes). Residues Pro710 to Arg814 are shown.

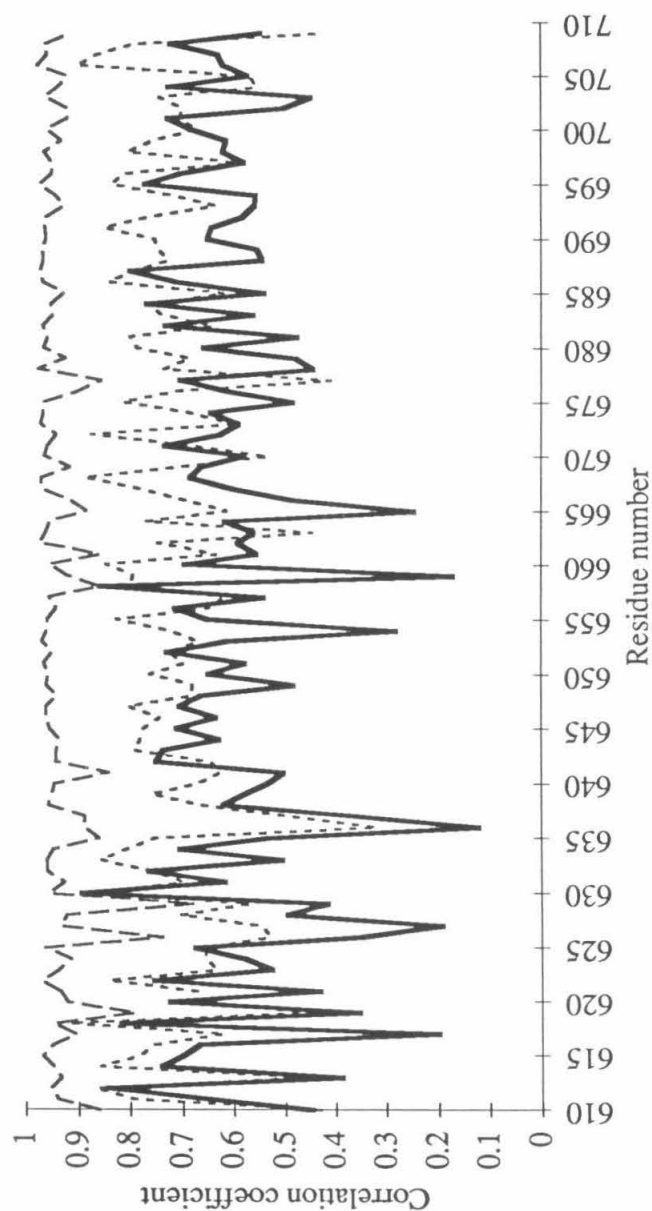


Figure 3-5A Plot of side chain real space correlation coefficient as a function of residue number. The real space fit of the final model was calculated with the Rs_fit utility in the program O, which expresses the quality of fit as a correlation coefficient. The real space fit was calculated separately for three electron density maps: a SIRAS map calculated to 2.6 Å (thick solid line), a 2.6 Å solvent flattened SIRAS (thin with short dashes), and the final [2Fo-Fc], ϕ_{calc} map (thin with long dashes). Residues Ile610 to Gln 709 are shown.

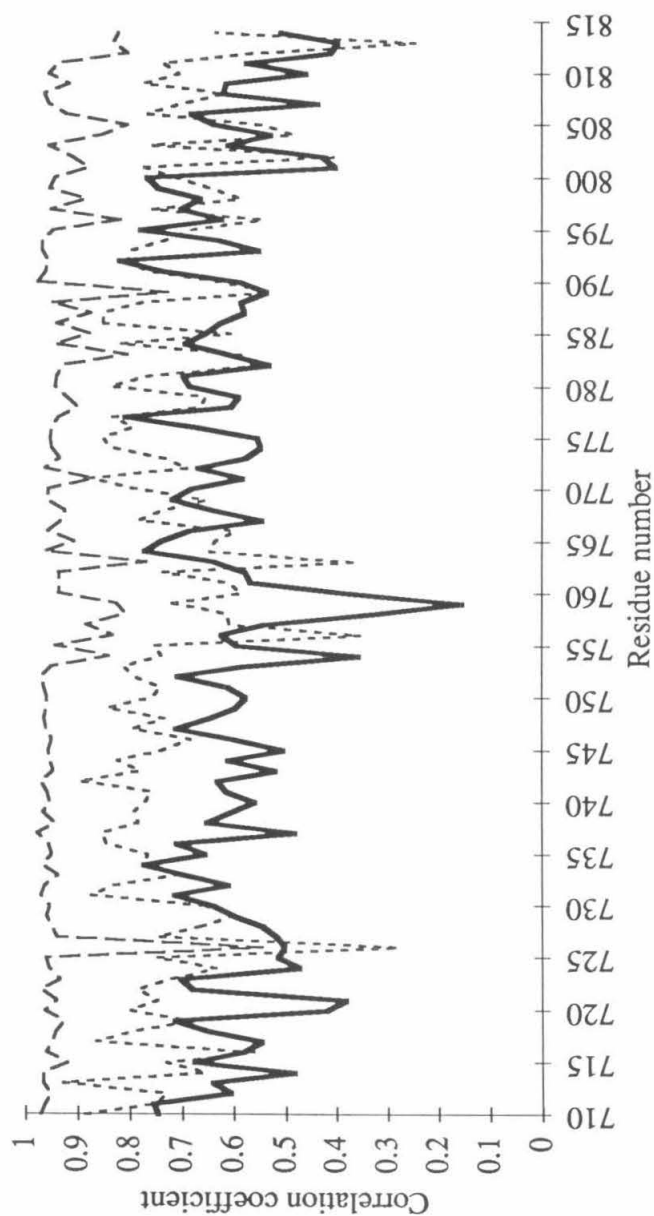


Figure 3-5B Plot of side chain real space correlation coefficient as a function of residue number. The real space fit was calculated separately for three electron density maps: a SIRAS map calculated to 2.6 Å (thick solid line), a 2.6 Å solvent flattened SIRAS (thin with short dashes), and the final $[2F_o - F_c]$, ϕ_{calc} map (thin with long dashes). Residues Pro710 to Arg814 are shown.

Structural mobility within the NgFn1,2 model was evaluated by plotting a backbone temperature factor versus residue number (Figure 3-6). X-PLOR used the individual B-values of backbone atoms to calculate a temperature factor for each residue. There are three regions that have elevated temperature factors in both domains: the A to B loop, the end of the C to C' loop (including the beginning of strand C'), and strand G.

Reliability of the NgFn1,2 structure. Several criteria used to judge the quality of structural models were evaluated for the NgFn1,2 structure. These criteria included R_{free} and R_{cryst} crystallographic residuals; r.m.s. deviations in bond lengths and angles; phi, psi, chi-1 and chi-2 dihedral angles; peptide bond planarity; carbon- α tetrahedral distortion; non-bonded and hydrogen bond energies; real space correlation coefficients for main and side chain atoms; backbone temperature factors and the deviation between model peptide carbonyl oxygen coordinates and those found in a database of well refined structures. The values derived from the NgFn1,2 model are consistent with a well refined high resolution structure.

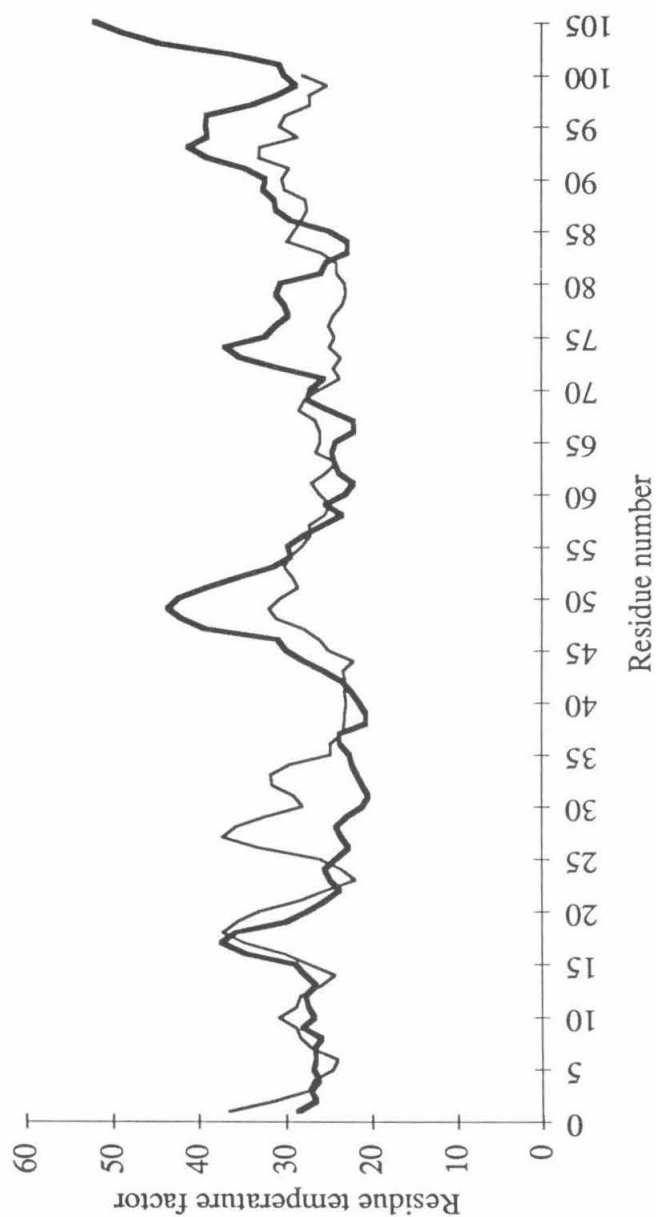


Figure 3-6 Plot of backbone temperature factor versus residue number. X-PLOR was used to calculate a temperature factor for each residue based on the individual B values of backbone atoms. DeltaGraph Professional for the Macintosh was then used to generate the plots. The NgFn1 domain starts with residue Ile610 (1 in figure) and is represented by the thin line. The NgFn2 domain starts with residue Pro710 (1 in figure) and is represented by the thick line.

References

1. Brünger, A.T. *X-PLOR (Version 3.1) Manual* (Yale University Press, New Haven, CT, 1992).
2. Powell, M.J.D. (1977). *Mathematical Programming* **12**, 241-254.
3. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
4. Luzzatti, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Cryst.* **5**, 802-810.
5. Leahy, D.J., Hendrickson, W.A., Aukhil, I. & Erickson, H.P. (1992). Structure of a Fibronectin Type III Domain from Tenascin Phased by MAD Analysis of the Selenomethionyl Protein. *Science* **258**, 987-991.
6. Bieber, A.J., Snow, P.M., Hortsch, M., Patel, N.H., Jacobs, J.R., Traquina, Z.R., Schilling, J. & Goodman, C.S. (1989). *Drosophila* Neuroglian: A member of the immunoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. *Cell* **59**, 447-460.
7. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for x-ray protein-structure refinement. *Acta Cryst.* **A47**, 392-400.
8. Weis, W.I., Drickamer, K. & Hendrickson, W.A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* **360**, 127-134.
9. Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R. & Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935.
10. Laskowski, R.A. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.
11. Jones, T.A. & Kjeldgaard, M. *Manual for O (version 5.9)* 1993).

12. Brändén, C.I. & Jones, T.A. (1990). Between objectivity and subjectivity. *Nature* **343**, 687-689.
13. Brünger, A.T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Cryst* **D49**, 24-36.
14. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.

Chapter 4

Identification of the Metal Ion Bound at
the Interface Between the First Two
Fibronectin Type III Repeats of *Drosophila*
Neuroglian

Introduction

An unexpected feature of the NgFn1,2 structure was the presence of a metal cation at the interdomain interface. The metal was first identified when strong spherical density was observed to lie within ≈ 2.2 Å of five potential oxygen ligands: the backbone carbonyl oxygens from Ile740 and Pro680, side chain oxygens from Ser679 and Asn743, and a water molecule (Chapter 2). The binding site has approximate square pyramidal coordination geometry with an average refined metal to ligand distance of 2.23 Å. This distance is significantly less than that expected for K^+ (2.7 to 2.8 Å) and between the metal to ligand distances commonly observed for Mg^{2+} (2.1 Å) and Na^+ or Ca^{2+} (2.4 Å)¹⁻³.

While the information present in the native structure does not permit an unambiguous identification of the bound metal ion, it does enable one to make an educated guess. Proteinaceous binding sites for divalent metal cations commonly utilize carboxylate oxygen atoms from aspartic and glutamic acid residues as metal ligands. These side chain groups are usually ionized under physiological conditions and contribute to the local neutralization of the metal ion's charge. The absence of any carboxyl ligands in the NgFn1,2 binding site suggests the site is specific for monovalent rather than divalent cations. Since the observed electron density is too strong for Li^+ and the average metal to ligand distance is too short for K^+ , the identity of a bound monovalent cation would most likely be Na^+ .

In order to further characterize the metal binding site and possibly identify the bound cation, data were collected from crystals soaked in solutions containing Na_2EDTA , $CaCl_2$, $NaCl$ or KCl . Difference Fourier analyses and ion occupancy refinement were used to monitor changes in the ion binding site.

Materials and Methods

Crystal Growth and Ion Soak Conditions

Native crystals of NgFn1,2 were grown as described in Chapter 2 and were soaked in solutions containing Na₂EDTA, CaCl₂, NaCl, or KCl overnight at 22°C. The CaCl₂ soak was preceded by an overnight soak in Na₂EDTA. The Na₂EDTA and CaCl₂ soak solutions consisted of 100 mM Hepes (sodium salt)/NaOH pH 7.5, 45% saturated Li₂SO₄ and either 20 mM Na₂EDTA or 10 mM CaCl₂. The NaCl and KCl soak solutions comprised 1.88 M (NH₄)₂SO₄, 100 mM Hepes (free acid)/NH₄OH pH 7.6 and either 10 mM NaCl or 10 mM KCl. Each crystal was transferred to a mounting buffer immediately before freezing. The mounting buffers were essentially the final soak conditions with 25% (v/v) glycerol as a cryoprotectant.

Data Collection and Processing

Single crystals were suspended in 1 mm diameter glass loops by a thin film of mounting buffer⁴ and were flash frozen using an LN₂ low temperature system (Molecular Structure Corporation). The average temperature of the N₂(g) coldstream during crystal freezing and data collection was approximately -165 °C. Diffraction data were collected with an R-Axis IIC x-ray detector mounted on a “Rotaflex” RU-200 Series rotating anode x-ray generator (Rigaku Corporation). The determination and refinement of crystal orientation as well as the indexing and integration of the raw data were performed with the R-Axis IIC data processing software. Profile fit intensities were scaled and merged into unique reflections with the ROTAVATA and AGROVATA programs from the CCP4 package (CCP4, The SERC (UK) Collaborative Computing Project No. 4. A Suite of Programs for Protein Crystallography distributed from Daresbury Laboratory, Warrington, WA44AD, UK (1979)) (Table 4-1).

Map and Occupancy Calculations

Data sets were scaled to each other using the RSTATS program (CCP4 package) which performs a non-linear least squares minimization of a linear scale factor and an exponential temperature factor. Difference maps, using coefficients ($F_{\text{soak}} - F_{\text{native}}$) or ($F_{\text{KCl}} - F_{\text{NaCl}}$), were calculated with the FFT program (CCP4 package) using phases derived from refined native structures. Maps were interpreted with the program O⁵ on IRIS 4D series work stations (Silicon Graphics Incorporated).

Occupancy refinement was carried out with the program X-PLOR⁶. The refined native coordinates, with the B value of the ion set to the average ligand B value, were used in all occupancy calculations.

Results and Discussion

Data Quality

Overall, the soak data are less reliable than the native data (as judged by R_{merge} values for data sets) and were collected to a significantly lower resolution (Table 4-1). This discrepancy in data quality is not surprising. The native data were collected at a synchrotron, which provides a very intense source of x-rays. A rotating anode x-ray generator, which provides significantly less x-ray flux than a synchrotron, was used during the collection of the soak data sets. The higher flux available at a synchrotron results in a better signal to noise ratio and allows the collection of data that would be too weak to collect using a rotating anode x-ray source. There was also a difference in the quality of the crystals from which the data were collected. Many of the crystals used in the soak experiments were smaller and older than the crystal used to collect the native data. These differences in available x-ray flux and crystal quality make it difficult to assess how much the ion soak conditions affected the crystalline lattice. While it is likely that the ion soaks

damaged the crystals, none of the soak conditions visibly disrupted the crystalline lattice or produced significant changes in the unit cell dimensions (Table 4-1).

NgFn1,2 Metal Binding Site is not Specific for Divalent Cations

The intensity of the ion peak in the native structure was comparable to that observed for well ordered water molecules (data not shown), suggesting that the ion was either a Na^+ or Mg^{2+} with relatively high occupancy or a Ca^{2+} with lower occupancy. In order to test the possibility that the bound ion was a Mg^{2+} , native crystals were soaked in 20 mM Na_2EDTA . If EDTA chelated the bound ion, one would expect to observe a strong negative peak at the ion position in $(F_{\text{Na}_2\text{EDTA}} - F_{\text{native}})$ difference Fourier maps. Difference Fourier analysis produced a strong positive peak instead of a strong negative peak. This suggested the site was being occupied by Na^+ ions (Table 4-2) which were present in the Na_2EDTA and the Hepes buffer used to make the soak solution. The refined occupancy of a Na^+ at this site is 100% (Table 4-3). These data strongly suggest that the metal binding site observed in the native crystals is not occupied by a Mg^{2+} ion.

The observed density for the ion in the native structure could be modeled as a Ca^{2+} with 47% occupancy (Table 4-3). Crystals were soaked in CaCl_2 in an attempt to either bring the proposed Ca^{2+} ion to full occupancy or to exchange Ca^{2+} for the bound ion. Ca^{2+} binding would generate a large positive peak in $(F_{\text{CaCl}_2} - F_{\text{native}})$ difference Fourier maps. These soak conditions produced only a small positive difference peak, suggesting that the protein was binding the Na^+ present in the soak solution rather than the Ca^{2+} (Table 4-2). This conclusion is supported by occupancy refinement calculations which estimate the occupancy of the ion to be 95% when modeled as a Na^+ (Table 4-3). Ca^{2+} is significantly more electron dense than Na^+ and thus modeling a fully occupied Ca^{2+} ion as a Na^+ ion should produce an occupancy greater than one.

The results of the Na₂EDTA and CaCl₂ soak experiments strongly suggest that NgFn1,2 does not preferentially bind divalent cations. In fact, these experiments demonstrated that the metal site could bind Na⁺ ions. As described earlier, an absence of carboxylate oxygen ligands and an observed metal to ligand distance of 2.23 Å suggest that a Na⁺ ion is bound in the native crystals.

NgFn1,2 metal binding site specifically binds Na⁺

The crystal used to solve the native structure was exposed to a relatively high concentration of Na⁺ ions. The synthetic mother liquor used to harvest and preserve the native crystal contained of 45% saturated (22 °C, ≈1 M) Li₂SO₄ and 100 mM Hepes (sodium salt)/NaOH pH 7.5. At first glance, the high Na⁺ ion concentration suggests that the ion might be bound nonspecifically. However, if NgFn1,2 nonspecifically bound monovalent cations, the approximately 20-fold higher concentration of Li⁺ would favor the binding of Li⁺ ions. The refined occupancy of a Li⁺ ion in the native structure is 400%, demonstrating that the majority of the bound ion is Na⁺. While Na⁺ does appear to be preferentially bound, the refined occupancy for a Na⁺ ion is only 77%, suggesting that Li⁺ can compete for binding when present at high concentrations.

The Na₂EDTA and CaCl₂ soak solutions also contained at least 100 mM Na⁺ because the sodium salt of Hepes was used as a buffer. Thus, one could argue that the binding of Na⁺ observed in these experiments was nonspecific. In order to further demonstrate that NgFn1,2 specifically binds Na⁺ ions, native crystals were soaked in a solution containing only 10 mM NaCl. The other components of this soak solution, (NH₄)₂SO₄ and Hepes (free acid)/NH₄OH, were not potential sources of monovalent cations. The NaCl soak data produced a positive difference peak (Table 4-2) at the ion position and a refined Na⁺ occupancy of 100%, confirming that Na⁺ binding is specific.

NgFn1,2 metal binding site does not bind K⁺

The existence of a metal binding site at the NgFn1,2 interdomain interface suggests that the relative orientation of the domains is stabilized by the binding of a metal ion, in this case Na⁺. The location of this site also raises the possibility that the relative orientation of the two domains could be altered by substituting a different ion for the Na⁺. Since there are not any carboxylate groups near the metal binding site, it is likely that the Na⁺ would be exchanged for another monovalent cation such as K⁺. The substitution of Na⁺ for K⁺ at an ion binding site has been observed to produce significant conformational changes in the enzyme dialkylglycine decarboxylase (DGD)³. This enzyme requires K⁺ ions for stability and activity and is inhibited by Na⁺ and Li⁺ ions. The bound monovalent cation is near the active site but is not close enough to play a direct role in catalysis. In the DGD structure, replacing a bound K⁺ with Na⁺ requires a gross change in the ion coordination geometry and results in conformational changes that extended far beyond the local environment of the ion binding site. By analogy, substitution of K⁺ for Na⁺ in the NgFn1,2 structure could produce significant changes in the relative orientation of the two FnIII domains. Thus, if both domains of NgFn1,2 contribute to the binding site for a receptor, neuroglian-receptor interactions could be modulated by the identity of the ion bound at the NgFn1,2 interdomain interface.

The ability of NgFn1,2 to bind K⁺ was tested by soaking native crystals in an (NH₄)₂SO₄ based synthetic mother liquor containing 10 mM KCl. This soak condition did not produce positive ($F_{\text{KCl}} - F_{\text{native}}$) difference map peaks at the ion position but rather strong negative peaks that enveloped the ion and the ligand oxygen provided by the Ser679 side chain. The observed negative difference density and the refined ion (Na⁺) occupancy of 42% demonstrate that K⁺ is not bound by the protein and that the KCl soak actually decreases the occupancy of the ion site. The negative density on the Ser679 side chain oxygen and the presence of nearby positive difference density suggests that the side chain

can adopt an alternate conformation in the absence of a bound ion. A similar positive density peak is present in native ($F_{\text{observed}} - F_{\text{calculated}}$) and ($2F_{\text{observed}} - F_{\text{calculated}}$), ϕ_{calc} difference maps. As mentioned previously (above and in chapter 2), the calculated occupancy for a Na^+ ion in the native structure is 77%. These data suggest that the metal binding site in the native crystal is either partially empty or the Ser679 side chain oxygen is not needed for Li^+ binding.

It is possible that crystallized NgFn1,2 cannot bind K^+ because the binding event requires a conformational change prohibited by crystal lattice contacts. Thus, the results of the KCl soak experiments do not prove that the protein in solution is incapable of binding K^+ . This possibility can be tested by assaying the ability of the solubilized protein to bind a radioisotope of K^+ or by solving the NgFn1,2 structure with crystals grown in the presence of K^+ .

Table 4-1 Data collection statistics.

DATA	Native	NaCl	Na ₂ EDTA	CaCl ₂	KCl
Number of crystals	1	1	1	1	1
X-ray source	CHESS	RA	RA	RA	RA
Data collection temperature (°C)	-185	-165	-165	-165	-165
Cell edge (Å)	241.8	240.8	241.3	242.1	242.0
Resolution (Å)	1.8	2.5	2.5	2.5	2.9
Number of measurements	355,936	121,756	138,876	112,112	92,566
Unique reflections	45,766	20,743	21,250	19,963	10,491
% Complete (Resolution Å)	99 (2.0)	98 (2.5)	99.6 (2.5)	93.1 (2.5)	99.9 (3.2)
R merge (Resolution Å)	0.050 (1.75)	0.089 (2.8)	0.070 (2.69)	0.087 (2.69)	0.120 (3.20)

$R_{\text{merge}} = \Sigma (|I - \langle I \rangle|) / (\Sigma \langle I \rangle)$, where I is the intensity of a reflection, and $\langle I \rangle$ is the mean intensity for the reflection; CHESS is the Cornell High Energy Synchrotron Source; RA is a "Rotaflex" RU-200 Series rotating anode x-ray generator from Rigaku Corporation.

Table 4-2 Difference map peaks at ion binding site.

<u>Map</u>	<u>position</u>	<u>difference peak value (σ_{map} multiple)</u>
F _{Na} -F _{Native}	ion	4.4
F _{EDTA} -F _{Native}	ion	4.8
F _{Ca} -F _{Native}	ion	2.7
F _K -F _{Native}	ion	-4.5
F _K -F _{Native}	Ser 679, O γ	-6.0
F _K -F _{Na}	ion	-7.6
F _K -F _{Na}	Ser 679, O γ	-8.2

Table 4-3 Refined values for ion occupancy.

<u>Soak</u>	<u>Ion</u>	<u>Resolution of data (Å)</u>	<u>Occupancy (%)</u>
None (Native)	Na ⁺	10.0 - 2.00	77
None (Native)	Na ⁺	10.0 - 2.80	74
None (Native)	Na ⁺	10.0 - 3.20	86
None (Native)	Ca ²⁺	10.0 - 2.00	47
None (Native)	Li ⁺	10.0 - 2.00	400
Na ₂ EDTA	Na ⁺	10.0 - 2.65	100
CaCl ₂	Na ⁺	10.0 - 2.65	95
NaCl	Na ⁺	10.0 - 2.80	104
KCl	Na ⁺	10.0 - 3.20	42

References

1. Yamashita, M.M., Wesson, L., Eisenman, G. & Eisenberg, D. (1990). Where metal ions bind in proteins. *Proc. Natl. Acad. Sci. USA* **87**, 5648-5652.
2. Glusker, J.P. (1991). Structural requirements of metal liganding to functional groups in proteins. *Adv. Prot. Chem.* **42**, 1-76.
3. Toney, M.D., Hohenester, E., Cowan, S.W. & Jansonius, J.N. (1993). Diallylglycine decarboxylase structure: bifunctional active site and alkali metal sites. *Science* **261**, 756-759.
4. Teng, T.-Y. (1990). Mounting of crystals for Macromolecular Crystallography in a Free-Standing Thin Film. *J. Appl. Cryst.* **23**, 387-391.
5. Jones, T.A. & Kjeldgaard, M. *Manual for O (version 5.9)* 1993).
6. Brünger, A.T. *X-PLOR (Version 3.1) Manual* (Yale University Press, New Haven, CT, 1992).

Chapter 5

Characterization of the Extracellular
Domains of *Drosophila* Neuroglial By
Electron Microscopy

Introduction

As described in Chapter 2, the structure of the first two fibronectin type III repeats from *Drosophila* neuroglian (NgFn1,2) was used to model the five tandem fibronectin type III repeats (NgFn1-5) in whole neuroglian. In the resulting model, the fibronectin type III (FnIII) domains pack end-to-end to form a 175 Å long linear rod with a two domain zig-zag repeat. If one assumes that neuroglian's six tandem Ig-like domains (NgIg1-6) also form a linear array, the dimensions of NgIg can be estimated using known structures. Pairs of tandem Ig-like domains from CD4¹⁻³ and CD2⁴ are 60-65 Å long, suggesting that NgIg is approximately 180-195 Å long. Given the above models, the entire extracellular portion of neuroglian (Ng), comprising six tandem Ig-like domains followed by five FnIII repeats, is predicted to be a long, thin rod that extends up to 370 Å from the cell surface.

While these models do give an estimate of the overall dimensions of linear NgFn1-5 and Ng molecules, they do not address questions concerning molecular flexibility and overall morphology. Of the ten interdomain interfaces between tandem domains in the extracellular portion of Ng, only the interface between the first two FnIII repeats, NgFn1 and NgFn2, has been structurally characterized. The remaining interdomain interfaces are potential sites of molecular flexibility. Electron micrographs of rotary shadowed N-CAM, L-CAM and I-CAM have demonstrated that each of these intercellular adhesion molecules forms a long rod with a single flexible bend^{5,6}. This flexibility may promote intercellular adhesion by enabling proteins to form interactions that would not be sterically possible with pairs of rigid rod-like molecules⁵. It has also been suggested that a high degree of flexibility would facilitate diffusion of the molecule's binding site(s) in the extracellular environment and thereby enhance binding kinetics⁶.

Electron micrographs of rotary shadowed NgFn1-5 and Ng were generated by Harold Erickson (Duke University) in order to view the conformations adopted by these molecules in solution. Hundreds of molecules, prepared under identical experimental conditions, can be viewed in a single electron micrograph. This allows a reliable estimation of the overall dimensions and flexibility of the protein. Rotary shadowed NgFn1-5 and Ng are thin rods with lengths of approximately 190 Å and 380 Å, respectively. These dimensions are consistent with the modeling of these proteins as linear arrays of tandem repeats aligned end to end. Ng and NgFn1-5 are observed to be highly flexible molecules with up to four and two bends, respectively. Neuroglian has been reported to mediate homophilic adhesion⁷(A. Bieber et al., unpublished results). However, no evidence of homophilic dimerization or aggregation was observed with NgFn1-5 or Ng, suggesting that the homophilic affinity of the molecules in solution is weak.

Materials and Methods

Protein Expression and Purification

Soluble forms of Ng and NgFn1-5 were expressed as secreted proteins in *Drosophila* S2 cells and were purified from the growth medium by immunoaffinity chromatography (chapter 2). The expression vectors and transfected cell lines used to produce Ng and NgFn1-5 were generated in the lab of Dr. Allan Bieber at Purdue University.

Electron Microscopy

Dr. Harold Erickson of Duke University prepared electron micrographs of rotary shadowed Ng and NgFn1-5 as described previously⁸. All photomicrographs of the observed images were generated at a total magnification of 150,000x.

Measurement of Images in Photomicrographs

Molecular dimensions were measured to the nearest 0.1 mm on photomicrographs using a 7X measuring magnifier (Bausch & Lomb). Straight segments of bent molecules were measured as shown in Figure 5-1. One nm was subtracted from the calculated length of segments containing a molecular terminus in order to compensate for the metal shell deposited during rotary shadowing. Measured molecules were not randomly selected, but rather chosen from a large population because they had clearly defined molecular borders and a specific shape (e.g., straight, one bend, etc.).

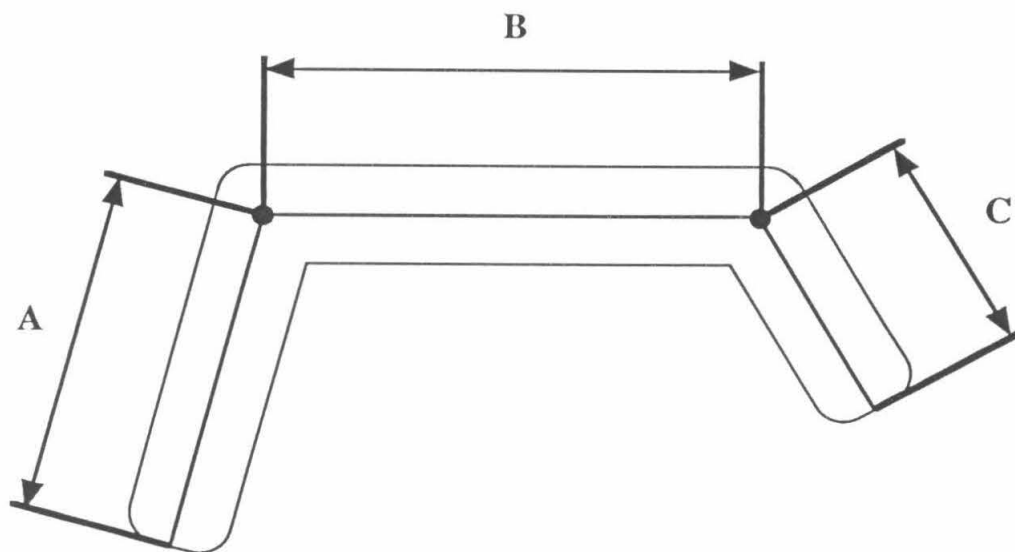


Figure 5-1 Measurement of segments from a bent molecule. The arrows show the distances that would be measured for molecular segments A, B and C of a molecule with two bends.

Estimation of the Number of Domains in NgFn1-5 Arms

The total length measured for straight NgFn1-5 proteins was assumed to be more accurate than that measured for bent molecules because it is difficult to measure bent regions accurately. The mean length of straight NgFn1-5 was divided by five to obtain the average translation per FnIII domain (38.6Å). The number of domains in the straight segments of bent NgFn1-5 molecules was determined by dividing segment length by 38.6 Å.

Results and Discussion

Electron Micrographs of the Five Neuroglian FnIII Repeats (NgFn1-5)

Straight molecules as well as molecules with one or two clearly defined bends could be identified in photomicrographs of rotary shadowed NgFn1-5. The large number of observed molecular conformations strongly suggests that NgFn1-5 is a highly flexible protein.

Straight NgFn1-5 molecules. NgFn1-5 can form straight rods with an average length of 193 Å (Table 5-1, Figure 5-2A). However, these molecules account for a small percentage (5% or less; data not shown) of the total molecular population, suggesting that the protein is highly flexible with only a few molecules having a straight conformation at any given time. Assuming the translation and rotation relating each pair of tandem domains in NgFn1-5 is approximately the same, the observed molecular length corresponds to a 38 Å translation per domain. These dimensions are consistent with a linear array of tandem FnIII repeats having a 37.7 Å translation per domain as modeled using the NgFn1,2 structure (chapter 2).

NgFn1-5 molecules with a single bend. Visual inspection identified two major populations of single-bend NgFn1-5 molecules (Figure 5-2B,C,D). These molecules have arms of two and three domains or arms of one and four domains (Tables 5-1 and 5-2). The angle between the arms of single-bend molecules was observed to vary from close to 180 ° (Figure 5-2C) to approximately 90° (Figure 5-2D), demonstrating that the bend regions are quite flexible.

The existence of two distinct populations of single-bend molecules implies that NgFn1-5 must contain at least two points of flexibility. It is not possible to determine from the photomicrographs alone which domains flank the bend regions because all five domains are homologous in size and shape. However, crystallographic and biochemical characterization of NgFn1,2 strongly suggests that the interface between the first two FnIII repeats is not flexible (chapter 2). Thus, the electron micrograph data is best explained by a model in which bending occurs between the fourth and fifth FnIII repeats (NgFn4 and NgFn5) and between either or both of the domain pairs NgFn2,3 and NgFn3,4.

NgFn1-5 molecules with two bends. Virtually all molecules with two clearly identifiable bends had a two domain middle segment and arms of one and two domains (Figure 5-2E,F). These images, as well as those of single bend molecules, are consistent with a model of NgFn1-5 capable of flexing between the second and third FnIII domains and between the fourth and fifth FnIII domains. Although a model with just two points of flexibility can account for all of the observed data, it is not possible to rule out bending at the NgFn3,4 interface. Molecules with a single bend at this interface cannot be distinguished from molecules which bend at the NgFn2,3 interface. Bending at these sites could be distinguished in two-bend proteins, but it is likely that molecules with bends at NgFn2,3 and NgFn3,4 or NgFn3,4 and NgFn4,5 are not clearly defined in the photomicrographs.

NgFn1-5 molecules with two bends can be divided into two groups: 'cis' proteins with both arms on the same side of the middle segment (Figure 5-2E) and 'trans' molecules with arms on opposite sides of the middle segment (Figure 5-2F). The existence of cis and trans molecules suggests that at least one of the two bend regions also has a large degree of rotational freedom.

Electron Micrographs of the Entire Extracellular Portion of Neuroglian

Ng, the extracellular portion of neuroglian, was found to be a long, thin and highly flexible protein.

The absence of straight Ng molecules. None of the images in the photomicrographs of Ng were clearly identified as a straight molecule. This apparent absence of straight proteins could be due to a high degree of molecular flexibility. If the protein contained several highly flexible bend regions, very few molecules would be expected to be in a straight conformation at any given moment in time. It is also possible that straight molecules are not observed because Ng has a structurally defined bend. Since NgFn1-5 is capable of adopting a straight conformation, this bend would have to occur within the tandem Ig-like domains or at the interface between the sixth Ig-like domain and the first FnIII repeat.

Ng molecules with bends. Measurement of Ng molecules containing 1, 2 or 3 defined bends yields a mean total molecular length of approximately 380 Å. Given the measured length of straight NgFn1-5 (Table 5-1), the six tandem Ig-like domains (NgIg1-6) are calculated to be approximately 190 Å in length, consistent with our modeling of these repeats as a linear tandem array (chapter 2).

Ng molecules with up to four distinct bends were identified in photomicrographs (Figure 5-3). Since only two bends were observed in NgFn1-5 molecules, at least two of

the four bends in Ng involve Ig-like domains. One of these bend regions must lie within the tandem Ig-like domains, with the other site possibly located at the interface between NgIg6 and NgFn1. Consistent with this suggestion is the observation that the neural cell adhesion molecule N-CAM is flexible at the interface between its five Ig-like domains and two FnIII repeats¹.

The Affinity of Ng-Mediated Homophilic Adhesion

No homophilic dimers or higher order aggregates were observed in electron micrographs of NgFn1-5 or Ng. The homophilic adhesion molecule N-CAM also fails to dimerize when prepared under similar conditions⁵. This suggests that N-CAM and neuroglial mediate homophilic adhesion through one or more relatively weak interactions that may require the proteins to assemble in large cell-surface arrays. This view is supported by experiments demonstrating that N-CAM homophilic binding is weak but highly cooperative⁹.

How Flexibility Might Affect Function

As proposed for other rod-like adhesion molecules with single flexible bends⁵, homophilic adhesion between opposing neuroglial molecules on moving or noncomplementary surfaces may require a high degree of molecular flexibility (Figure 5-4A). Also, numerous bend points may be required to make multiple binding sites on neuroglial simultaneously accessible to different receptors. A high degree of flexibility combined with multiple binding sites could allow neuroglial to interact with several different types of receptors independently or simultaneously and might result in the formation of a complex web of intermolecular interactions (Figure 5-4B).

Table 5-1 Molecular measurements from photomicrographs.

<u>molecule</u>	<u>observed shape</u>	<u>number of observations</u>	<u>mean length of segment (Å)</u>		<u>total length (Å)</u>
			<u>arm A</u>	<u>middle</u>	<u>arm B</u>
NgFn	straight	46	193 ± 15	NA	NA
	1 bend, 4&1 domains	15	151 ± 13	NA	53 ± 8
	1 bend, 3&2 domains	15	127 ± 12	NA	87 ± 9
	2-bend, trans	15	81 ± 7	87 ± 9	49 ± 6
	2-bend, cis	15	79 ± 6	87 ± 8	50 ± 8
Ng	1 and 2 bend	117	NA	NA	NA
					392 ± 24

NA = not applicable.

Table 5-2 Number of FnIII repeats in arms of bent NgFn molecules.

<u>molecule shape</u>	<u>number of domains</u>		
	<u>arm A</u>	<u>middle</u>	<u>arm B</u>
1 bend, 4&1 domains	3.9	NA	1.4
1 bend, 3&2 domains	3.3	NA	2.3
2-bend, trans	2.1	2.3	1.3
2-bend, cis	2.0	2.3	1.3

NA = not applicable.

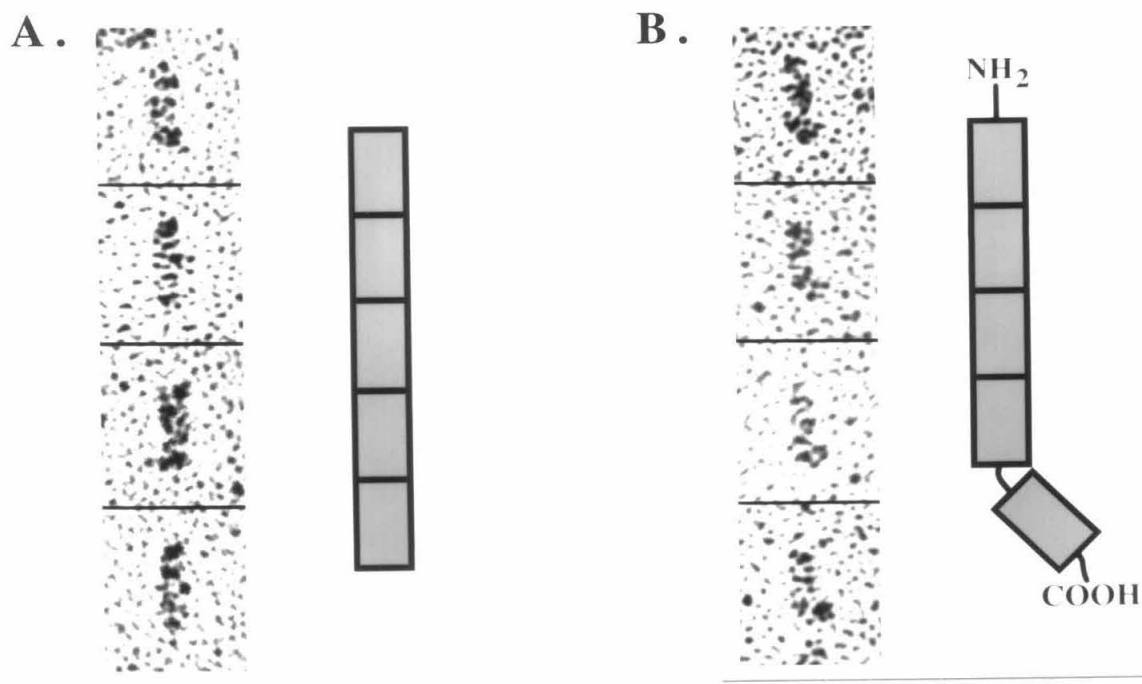
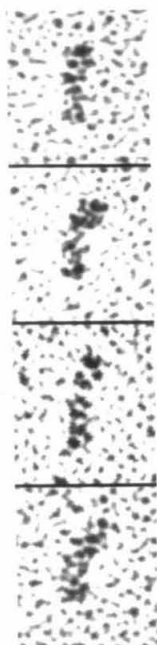
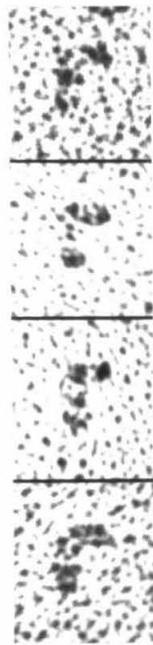


Figure 5-2 Representative photomicrographs of rotary shadowed molecules juxtaposed with schematic models of the observed proteins. Four images are shown of each observed molecular shape: (A) straight NgFn1-5 molecules, (B) single-bend NgFn1-5 molecules with arms of 1 and 4 domains, (C) single-bend NgFn1-5 molecules with a large angle between arms of 2 and 3 domains, (D) single-bend NgFn1-5 molecules with a small angle between arms of 2 and 3 domains, (E) two-bend NgFn1-5 molecules with a 'cis' conformation, (F) two-bend NgFn1-5 molecules with a 'trans' conformation.

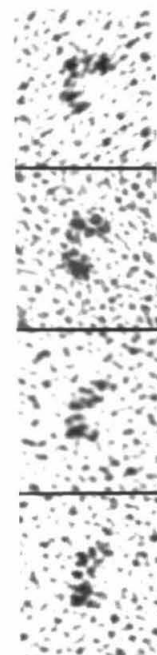
C.



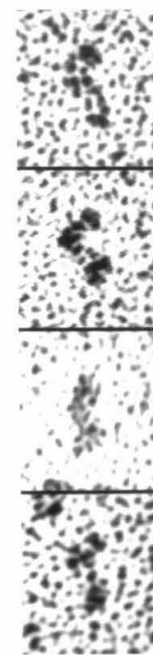
D.



E.

 NH_2 COOH

F.

 NH_2 COOH

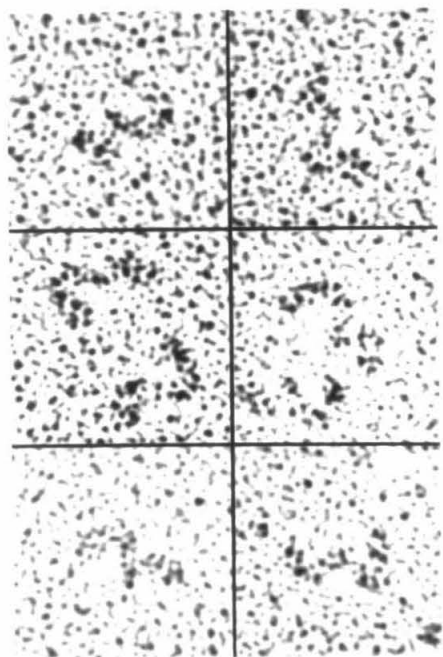
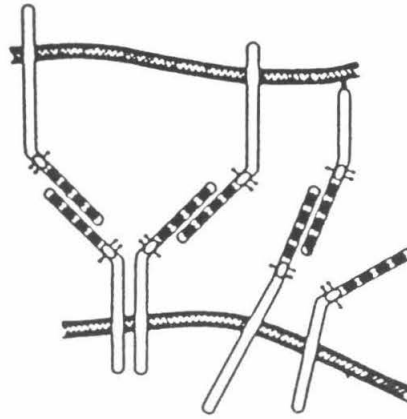


Figure 5-3 Photomicrograph images of Ng molecules with at least four bends.

A.



B.

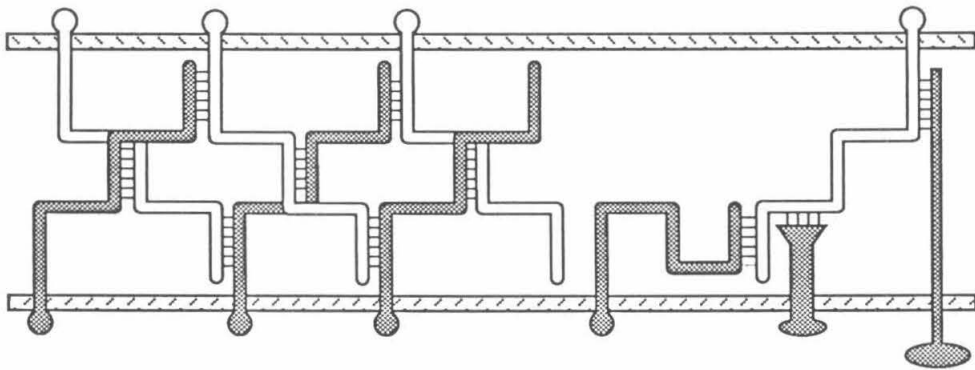


Figure 5-4 Schematic models of N-CAM- and neuroglial-mediated intercellular adhesion. (A) Arrays of N-CAM molecules on one cell interact with similar arrays on apposing cells. The flexible hinge in N-CAM may allow specific intermolecular contacts to be made even when cell surfaces are in motion or are not complementary in shape. (From Becker, J.W., Erickson, H.P., Hoffman, S., Cunningham, B.A. & Edelman, G.M. (1989). Topology of cell adhesion molecules. *Proc. Natl. Acad. Sci. USA* **86**, 1088-1092.). (B) Multiple points of flexibility could enable neuroglial to simultaneously interact with multiple receptors. Neuroglial molecules are schematized as long, thin, transmembrane proteins with four bends. Homophilic adhesion (left) and heterophilic adhesion (right) are depicted although only homophilic binding has actually been demonstrated.

References

1. Ryu, S.-E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R.W. & Hendrickson, W.A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* **348**, 419-426.
2. Wang, J., Yan, Y., Garrett, T.P.J., Liu, J., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L. & Harrison, S.C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* **348**, 411-419.
3. Brady, R.L., Dodson, E.J., Dodson, G.G., Lange, G., Davis, S.J., Williams, A.F. & Barclay, A.N. (1993). Crystal Structure of Domains 3 and 4 of Rat CD4: Relation to the NH₂-Terminal Domains. *Science* **260**, 979-983.
4. Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K. & Stuart, D.I. (1992). Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature* **360**, 232-239.
5. Becker, J.W., Erickson, H.P., Hoffman, S., Cunningham, B.A. & Edelman, G.M. (1989). Topology of cell adhesion molecules. *Proc. Natl. Acad. Sci. USA* **86**, 1088-1092.
6. Staunton, D.E., Dustin, M.L., Erickson, H.P. & Springer, T.A. (1990). The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. *Cell* **61**, 243-254.
7. Grenningloh, G., Bieber, A.J., Rehm, E.J., Snow, P.M., Traquina, Z.R., Hortsch, M., Patel, N.H. & Goodman, C.S. (1990). Molecular genetics of neuronal recognition in *Drosophila*: evolution and function of the immunoglobulin superfamily cell adhesion molecules. *Cold Spring Harbor Symposia on Quantitative Biology* **55**, 327-340.
8. Fowler, W.E. & Erickson, H.P. (1979). Trinodular structure of fibrinogen: confirmation by both shadowing and negative stain electron microscopy. *J. Mol. Biol.* **134**, 241-249.
9. Hoffman, S., Sorkin, B.C., White, P.C., Brackenbury, R., Mailhammer, R., Rutishauser, U., Cunningham, B.A. & Edelman, G.M. (1982). Chemical characterization of a neural cell-adhesion molecule purified from embryonic brain membranes. *J. Biol. Chem.* **257**, 7720-7729.

Appendix A

**Crystallization and Stoichiometry of
Binding of a Complex between a Rat
Intestinal Fc Receptor and Fc**

Crystallization and Stoichiometry of Binding of a Complex between a Rat Intestinal Fc Receptor and Fc

Andrew H. Huber, Robert F. Kelley, Louis N. Gastinel and Pamela J. Bjorkman

CRYSTALLIZATION NOTES

Crystallization and Stoichiometry of Binding of a Complex between a Rat Intestinal Fc Receptor and Fc**Andrew H. Huber¹, Robert F. Kelley², Louis N. Gastinel¹ and Pamela J. Bjorkman†**¹*Division of Biology and Howard Hughes Medical Institute
California Institute of Technology, Pasadena, CA 91125, U.S.A.*²*Department of Protein Engineering, Genentech, Inc.
460 Point San Bruno Blvd, South San Francisco, CA 94080, U.S.A.**(Received 17 November 1992; accepted 24 November 1992)*

Fc receptors expressed in the gut of newborn rodents bind to maternal immunoglobulin in milk at pH 6.5, and transport it to the bloodstream of the neonate, where it dissociates at pH 7.4. The rat intestinal Fc receptor (FcRn) consists of a heavy chain, with significant sequence similarity to the heavy chain of class I MHC molecules, complexed to the class I light chain, β 2-microglobulin. Although FcRn is predicted to contain a groove analogous to that which serves as the MHC peptide-binding site, the immunoglobulin ligand of FcRn is a macromolecule instead of a peptide. We have expressed and crystallized a secreted form of FcRn, and here report the crystallization of a complex between FcRn and its Fc ligand. Isolated FcRn–Fc complexes crystallize in space group *I*222 or *I*2₁2₁2₁ with unit cell dimensions $a=125$ Å, $b=152$ Å and $c=216$ Å. The crystals diffract to 5.5 Å resolution with anisotropic diffraction to 3.5 Å. Data collection from cryopreserved crystals may allow the resolution limit to be extended, since the major reason for the poor resolution appears to be radiation decay. Even a low-resolution view of how FcRn binds Fc would be of interest to see if the binding site corresponds to the functional part of an MHC molecule. Since the structure of Fc is known, and a structure determination of FcRn is underway, it may be possible to locate the Fc binding site on FcRn at low resolution. As an initial characterization of the FcRn–Fc mode of interaction, and to facilitate the structure determination, we have determined the stoichiometry of binding of FcRn to Fc. We show that two FcRn molecules bind per Fc, as determined by analysis of gels of washed crystals, a column binding assay, and isothermal titration calorimetry.

Keywords: Fc receptor; stoichiometry of binding; isothermal titration calorimetry; protein crystallization; histocompatibility molecules

The transmission of maternal immunoglobulin G (IgG₊) to fetal or neonatal mammals provides an important defense to the newborn before its immune response is fully functional. Suckling rats passively acquire immunity by transporting IgG from milk across the intestine and into the circulation (Culbertson, 1938). Receptors on epithelial brush borders bind to the Fc portion of IgG at the

acidic pH of the gut lumen (6.0 to 6.5), and the IgG–receptor complex is transcytosed to the basolateral membrane, where the IgG dissociates from the receptor at the slightly basic pH of the blood (7.4) (Rodewald, 1976). Fc receptors (FcRn, neonatal) have been affinity-purified from rat intestinal epithelial cell brush borders, and shown to be heterodimers consisting of two polypeptide chains of relative molecular masses of 45,000–53,000 and 14,000 (Simister & Rees, 1985). The FcRn heavy chain is predicted to consist of three extracellular domains, a single transmembrane spanning region, and a cytoplasmic tail. The extracellular domains have significant sequence similarity to the corresponding domains of major histocompatibility complex (MHC) class I molecules (Simister & Mostov, 1989*a,b*). This sequence similarity implies a

† Author to whom all correspondence should be addressed.

‡ Abbreviations used: IgG, immunoglobulin G; Fc, common or effector region of immunoglobulins; FcRn, neonatal Fc receptor; MHC, major histocompatibility region; FPLC, fast protein liquid chromatography; Fab, antigen-binding region of immunoglobulins; C_H3, immunoglobulin heavy chain constant region 3.

structural homology that is further reinforced by the finding that the FcRn small subunit is $\beta 2$ -microglobulin (Simister & Mostov, 1985), the same light chain as employed by class I MHC molecules. Although the two heterodimers appear to be structurally related, their physiological ligands and functions are very different. MHC molecules bind intracellularly processed peptides and display them on the cell surface for recognition by T cells (Townsend & Bodmer, 1989), while the FcRn binds and transports IgG, a large macromolecule. Crystal structures of class I molecules reveal a deep groove located between two α -helices that has a shape ideally suited for their peptide-binding function (Bjorkman *et al.*, 1987*a,b*; Garrett *et al.*, 1989; Bjorkman & Parham, 1990; Madden *et al.*, 1991; Fremont *et al.*, 1992). The conservation of critical residues that form the class I groove suggests that this feature is preserved in FcRn (Simister & Mostov, 1989*a,b*), but the mode of binding to Fc is unknown. The structural similarity between this Fc receptor and class I MHC molecules affords an opportunity to study how evolution has used this structural motif for different purposes. As a first step in a structural analysis of FcRn and its mode of binding to IgG, we have prepared and crystallized complexes of FcRn bound to Fc. The stoichiometry of the FcRn-Fc interaction has been analyzed to facilitate the solution of the crystal structure, and to obtain an initial characterization of the binding of FcRn and its ligand.

(a) Isolation and crystallization of FcRn-Fc complexes

A soluble form of FcRn, consisting of the three extracellular domains of the heavy chain complexed to rat $\beta 2$ -microglobulin, was expressed in Chinese hamster ovary (CHO) cells (Gastinel *et al.*, 1992). The heterodimer was purified through use of its pH dependent binding to immobilized IgG (Gastinel *et al.*, 1992). Complexes of FcRn and rat Fc (Jackson ImmunoResearch) were prepared by overnight incubation (4°C) at pH 6.3 of a protein solution at a molar ratio of approximately two FcRn molecules per Fc. Unbound molecules were separated from the complex by gel filtration on a HiLoad™ 26/60 Superdex™ 200 prep grade FPLC column, and the peak corresponding to the complex was concentrated to 15 mg/ml. At a 2:1 molar ratio of FcRn to Fc, nearly all of the Fc protein appeared to migrate as a complex (data not shown), but unbound FcRn could not be fully separated from the complex peak, hindering the determination of the exact stoichiometry of binding. Crystals of the complex were grown by the micro-vapor diffusion method (McPherson, 1982), and initial crystallization conditions were obtained from a factorial trial screen (Jancarik & Kim, 1991). Large single crystals were grown from protein solutions (15 mg/ml, 10 mM-piperazine-*N,N'*-bis(2-ethane sulphonic acid pH 6.4, 0.05% NaN₃) in 0.67 M-sodium tartarate, 100 mM-citrate (pH 5.6), 200 mM-MgCl₂ (Fig. 1).

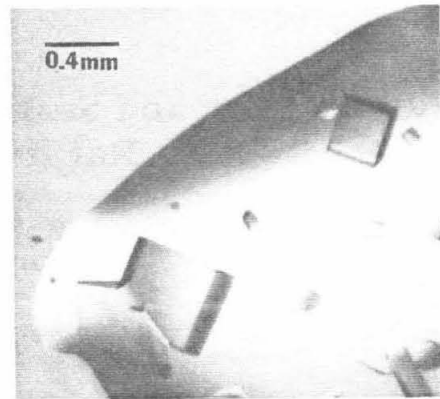


Figure 1. FcRn-Fc crystals. Length of crystals indicated by the bar.

SDS-PAGE analysis of washed crystals show bands corresponding to the FcRn heavy and light chains and Fc (Fig. 2).

The space group and cell dimensions of the crystals were determined by precession photo-

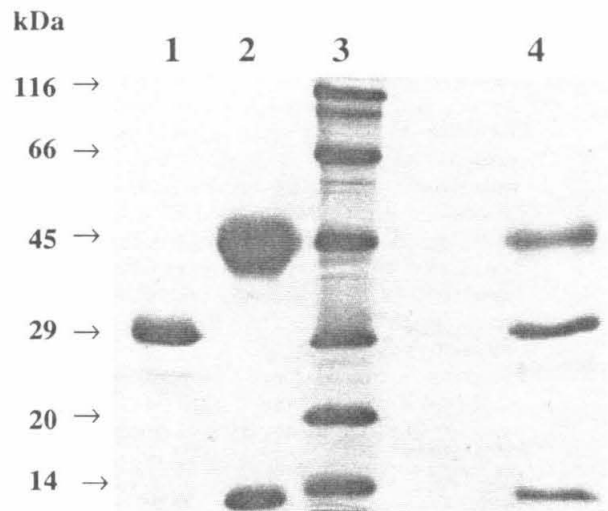


Figure 2. SDS/15% PAGE demonstration that washed crystals contain both Fc and FcRn proteins. Harvested crystals were washed in 0.75 M-NaCl, 100 mM-acetate (pH 4.6) and dissolved in reducing gel sample buffer. Lane 1, 3 μ g purified Fc (Jackson ImmunoResearch); lane 2, 25 μ g purified FcRn; lane 3, molecular weight standards; lane 4, washed crystals. Densitometry of two gels: four different known quantities of purified FcRn and three quantities of Fc were run on each SDS/15% PAGE gel. Gels were stained with Coomassie Brilliant Blue and individual lanes were densitometered using a Computing Densitometer (Molecular Dynamics) and the GELTOOL program. The measured density of each standard was plotted versus μ g of protein and used to evaluate the amount of FcRn and Fc present in three washed crystal samples. Assuming relative molecular masses of 45,000 for the FcRn heavy chain, 13,000 for the FcRn light chain, 60,000 for Fc (a dimer of two 30,000 polypeptide chains), the molar ratios (FcRn-Fc) for three individual lanes of washed crystals are 1.8, 1.9 and 1.6.

graphy and data collected on a Siemens proportional multiwire area detector mounted on a Siemens rotating anode (200 μm focal cup, 50 kV, 90 mA). Data were reduced with the XENGEN package of programs (Howard *et al.*, 1987). The space group of the crystals is $I222$ or $I2_12_12_1$ with unit cell dimensions $a=125$ Å, $b=152$ Å and $c=216$ Å (1 Å=0.1 nm). Based on average volume to mass ratios (V_m) of protein crystals (Matthews, 1968) and assuming a 2:1 molar ratio of FcRn molecules to Fc (see below), the asymmetric unit of the crystal is estimated to contain one complex and 57% solvent. Because the detector data were weak beyond 6 Å, the practical resolution limit of the crystals using a rotating anode is 6 Å. Preliminary diffraction photographs of complex crystals taken at the Stanford Synchrotron Radiation Laboratories ($\lambda=1.08$ Å) show anisotropic diffraction along the direction of the a -axis to 3.5 Å, with diffraction to 5.5 Å in the directions of the other two axes. We have obtained crystals of FcRn alone that diffract to 2.2 Å resolution (L.N.G., M. Blum & P.J.B., unpublished results), and a crystal structure determination is in progress. Using the FcRn structure and the known structure of Fc (Deisenhofer, 1981), it should be possible to locate the region of FcRn involved in binding to Fc even in a low-

resolution electron density map. To facilitate interpretation of the complex structure at low resolution, we determined the stoichiometry of FcRn-Fc binding.

(b) *Demonstration that two FcRn molecules can bind to a single Fc*

Densitometry of three individual gel lanes of washed crystals suggests a molar ratio of two FcRn molecules per Fc (see legend to Fig. 2). In order to verify these data, we devised a column binding assay to determine whether more than one FcRn molecule can bind per Fc.

Soluble FcRn was covalently coupled to CNBr-activated Sepharose 4B (Sigma). Immobilized FcRn bound IgG and Fc in the same pH-dependent manner as soluble FcRn and cell-surface FcRn (Simister & Rees, 1985), allowing execution of an experiment to determine whether two FcRn molecules can simultaneously bind to one Fc fragment (Fig. 3A). Purified rat Fc was bound to the column at pH 6.0, followed by 12 column volumes of pH 6.0 wash buffer (50 mM-phosphate (pH 6.0), 250 mM-NaCl). Soluble FcRn was then passed over the column, and the column was washed with an additional 20 column volumes of buffer. Proteins

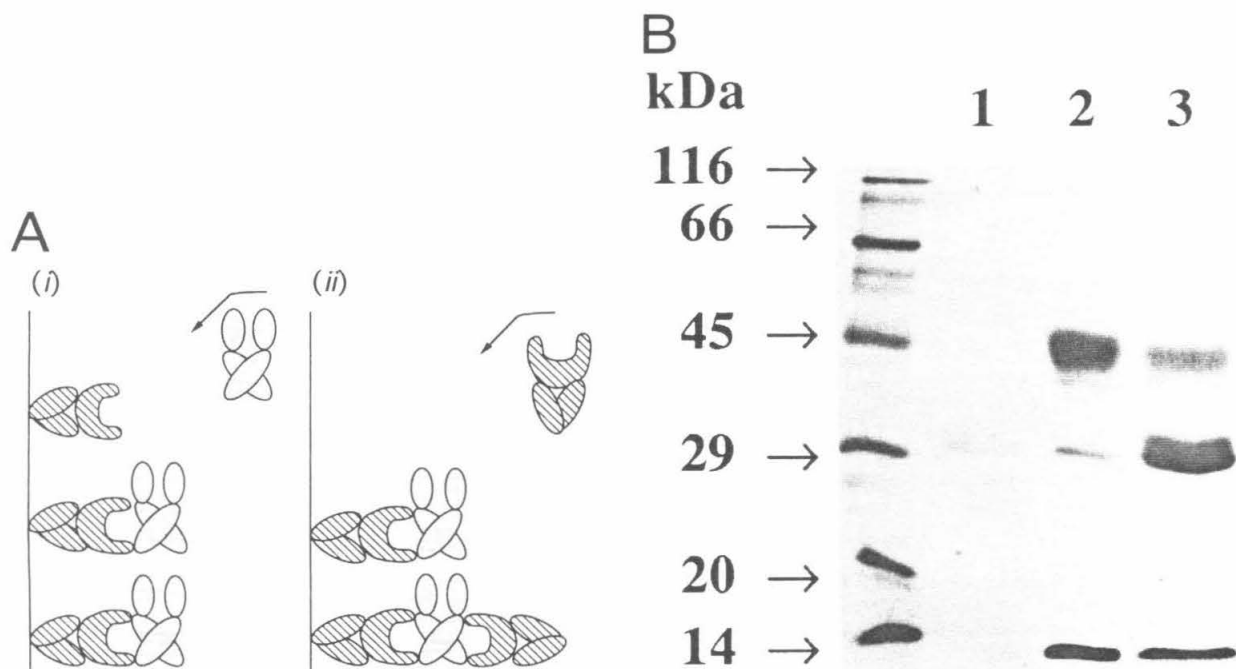


Figure 3. Demonstration that a single Fc can bind two FcRn molecules. A. Column experiment described schematically. (i) Purified FcRn (18 mg; represented schematically by hatched symbols) was covalently coupled to 2.5 ml CNBr-activated Sepharose 4B (Sigma). Samples of rat Fc (light shading) were passed over the column in 50 mM-phosphate (pH 6.0), 250 mM-NaCl, and the column was washed with 12 column volumes of the same pH 6.0 buffer. (ii) Purified FcRn (hatched symbols) was then passed over the column in 50 mM-phosphate (pH 6.0), 250 mM-NaCl, and the column was washed with 20 column volumes of the pH 6.0 buffer. Bound Fc and FcRn were eluted with 50 mM-phosphate (pH 8.0), 250 mM-NaCl, and the eluted fractions were analysed by SDS/17.5% PAGE under reducing conditions. B. lane 1, 28 μl of 1.0 ml fraction of Fc flowthrough; lane 2, 20 μl of 1.0 ml fraction of soluble FcRn flowthrough; lane 3, 53 μl of 1.0 ml fraction of protein eluted at pH 8.0. Because FcRn does not bind to the FcRn column in the absence of Fc (data not shown), any FcRn protein eluted from the column must be bound to an Fc molecule that is also binding immobilized FcRn.

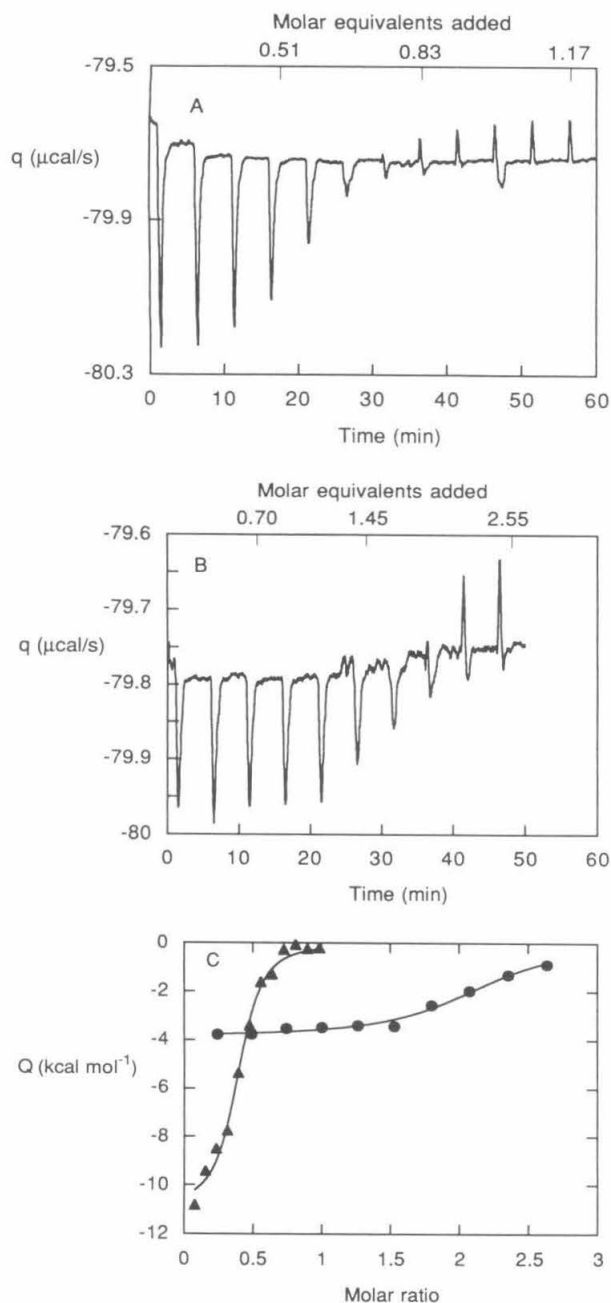


Figure 4. Calorimetric titration of FcRn and a human IgG1 monoclonal antibody (Carter *et al.*, 1992). Isothermal titration calorimetry was performed on a Microcal, Inc. OMEGA calorimeter at 25.5°C. Concentrated solutions of IgG and soluble FcRn in 20 mM-2-(*N*-morpholine) ethane sulfonic acid (pH 6.5), 0.1 M-NaCl were sterile filtered by passage through 0.22 µm Millipore Millex-GV filters. Protein concentrations were determined by absorbance measurements using an ϵ_{280} of 225 mm⁻¹ cm⁻¹ for the IgG and 84 mm⁻¹ cm⁻¹ for FcRn. The IgG and FcRn solutions were diluted to 10 µM (IgG) or 20 µM (FcRn) and one of the proteins was placed in the calorimeter cell (1.3943 ml). For these protein concentrations, the “*c*” value (Wiseman *et al.*, 1989) is between 10 and 100, and thus the parameters *n*, *K_a* and ΔH can be determined in a

directly and indirectly bound by the immobilized FcRn were eluted at pH 8.0, and the column flowthroughs and eluates were examined by SDS-PAGE (Fig. 3B). In the absence of Fc, FcRn does not bind to the FcRn column (data not shown). Since soluble FcRn could only be retained on the column by binding to an Fc molecule that was also bound by an immobilized FcRn, the presence of FcRn in the eluate indicates that more than one FcRn molecule can bind to a single Fc.

(c) *Isothermal titration calorimetry to measure stoichiometry of binding in solution*

Isothermal titration calorimetry may be used to determine the independent thermodynamic parameters (*K_a* and ΔH) as well as the number of binding sites (*n*) governing a protein–ligand interaction, provided that there is a measurable ΔH , and *K_a* is in the range of $10^3 < K_a < 10^9$ M⁻¹ (Wiseman *et al.*, 1989). Aliquots of a ligand protein are added to a solution in the calorimeter cell containing a fixed concentration of the binding protein, and the heat of reaction is measured until there is no further enthalpic change. If the stoichiometry is different from 1:1, the values of ΔH and *n* will differ depending on which component is in the calorimeter cell and which is being titrated. Since we can perform the titration experiment with either FcRn or its IgG ligand in the calorimeter cell, we have used this method to verify that the FcRn–Fc interaction has a 2:1 stoichiometry.

single experiment. The injection syringe (250 µl) was filled with a concentrated solution (0.154 mM) of the other component, mounted in the calorimeter, and allowed to come to thermal equilibrium at 25.5°C with stirring at 400 revs/min. A series of injections spaced 4 min apart were made, and the observed heat pulses were integrated using software supplied by the manufacturer. Heats of dilution measured in separate experiments by injection of IgG or FcRn into buffer were subtracted from the integrated values yielding the heats of binding. The thermodynamic parameters *n*, *K_a* and ΔH were determined by non-linear regression analysis of the binding isotherm. A. Calorimetric titration of FcRn with IgG at 25.5°C. Heat pulses observed for 12 injections (14 µl each) of a solution of 0.154 mM-IgG into a solution of 20 µM-FcRn. Injections were made every 4 min. B. Calorimetric titration of IgG with FcRn at 25.5°C. Heat pulses observed for 10 injections (22 µl each) of a solution of 0.154 mM-FcRn into a solution of 10 µM-IgG. Injections were made every 5 min. C. Stoichiometry of IgG–FcRn interaction. Binding isotherms observed by calorimetry for titration of FcRn with IgG (triangles) or for titration of IgG with FcRn (circles) are plotted *versus* the molar ratio of titrant to binding protein obtained at the end of the injection. The solid lines are the result of fitting the data to a single class of non-interacting binding site(s) model. For titration of FcRn with IgG, this analysis yielded the thermodynamic parameters: *n*=0.38 (±0.01), *K_a*=2.3 (±0.5) × 10⁶ M⁻¹, and ΔH =−11.0 (±0.4) kcal/mol. Titration of IgG with FcRn yielded the thermodynamic parameters: *n*=2.05 (±0.04), *K_a*=1.7 (±0.4) × 10⁶ M⁻¹, and ΔH =−3.9 (±0.1) kcal/mol.

Calorimetric measurements of the interaction of a human IgG1 monoclonal antibody (Carter *et al.*, 1992) with soluble FcRn are shown in Figure 4A for titration of FcRn with IgG, and in Figure 4B for the opposite experimental configuration. Non-linear regression analyses of the data using a single class of non-interacting binding site(s) model revealed a binding constant (K_a) of $2 \times 10^6 \text{ M}^{-1}$ ($K_D = 0.5 \text{ }\mu\text{M}$) for both experiments. However, the derived values of ΔH and n differed depending on which protein was being titrated. A larger value of n ($n = 2.05$) and smaller value of ΔH (-3.9 kcal/mol $1 \text{ cal} = 4.184 \text{ J}$) were obtained for titration of IgG with FcRn, as compared with titration of FcRn with IgG ($n = 0.38$, $\Delta H = -11.0 \text{ kcal/mol}$; Fig. 4A to C). The dependence of the values of ΔH and n upon which protein is being titrated suggests that the IgG–FcRn stoichiometry is different than 1:1, and both values of n predict that two FcRn molecules bind to a single IgG. The good fit of the data to a single class of non-interacting binding site(s) model suggests that more complicated models involving two sites of differing affinity, or cooperativity between the two sites, can be eliminated.

(d) Discussion

We have begun a structural investigation of a rat Fc receptor (FcRn) with an intriguing structural similarity to class I MHC molecules. The predominant feature in the crystal structures of class I MHC proteins is a groove located between two long α -helices, which is the binding site for peptides derived from self and foreign antigens (Bjorkman *et al.*, 1987a,b; Garrett *et al.*, 1989; Bjorkman & Parham, 1990; Madden *et al.*, 1991; Fremont *et al.*, 1992). Although FcRn is predicted to share the helices and β -sheet that form the MHC peptide-binding site (Simister & Mostov, 1989a), the immunoglobulin ligand of FcRn is a macromolecule instead of a peptide, and it is not obvious how the Fc portion of IgG could fit into a groove of similar dimensions as the class I MHC peptide binding site. Comparison with class I MHC molecules suggests two other candidate regions for the Fc-binding site (Simister & Mostov, 1989a); the counterparts of either the class I binding sites for T-cell receptor, or for CD8, a co-receptor on cytotoxic T cells. The hypervariable regions of the Fab-like T-cell receptor have been suggested to contact the two α -helices of the MHC molecule, necessitating simultaneous contact with a peptide occupying the site (Chothia *et al.*, 1988; Davis & Bjorkman, 1988; Claverie *et al.*, 1989), and CD8 interacts with the membrane proximal portion of the class I α_3 domain (Connolly *et al.*, 1988; Salter *et al.*, 1989). As an initial step in comparing the interaction between FcRn and its immunoglobulin ligand with the recognition of class I MHC–peptide complexes by immunoglobulin superfamily members, we have crystallized an FcRn–Fc complex. The crystals are extremely radiation sensitive, and reflections are only observed to low ($3.5 \text{ }\text{\AA}$ to $5.5 \text{ }\text{\AA}$) resolution at room

temperature. Using a high-intensity beam at a synchrotron facility, it is possible that higher resolution data can be collected from cryopreserved crystals (Hope, 1990), which have an essentially infinite lifetime in the X-ray beam. In the case of the FcRn crystals, we have observed a dramatic increase in resolution by collecting data at -165°C at the Cornell High Energy Synchrotron Source (CHESS). Under these conditions, data were collected to $2.2 \text{ }\text{\AA}$, as compared to a maximum resolution $\sim 3.5 \text{ }\text{\AA}$ at room temperature (L.N.G., M.L. Blum & P.J.B., unpublished results). The conditions for preparation of cryopreserved crystals of the FcRn–Fc complex are being explored. Ultimately, results of the structure determination of FcRn alone (Gastinel *et al.*, 1992) coupled with the known Fc crystal structure (Deisenhofer, 1981) should allow identification of the portions of FcRn and Fc involved in binding even at low resolution.

Because the Fc portion of an IgG is a chemical dimer, it was possible that two FcRn molecules bind per Fc, analogous to the observation of a 2:1 stoichiometry in the structure of the complex of a fragment of protein A with Fc (Deisenhofer, 1981). A 1:1 stoichiometry of binding was possible in the event that the binding of one FcRn sterically hinders a second from binding, or the binding site on Fc is represented only once in its structure (for example, the binding site is formed at the bottom of the Fc molecule spanning the two C_H3 domains). Densitometry of gels of washed crystals suggest that the crystallized complex contains two FcRn molecules per Fc. To verify that more than one FcRn can bind per Fc in solution, we used a column-binding assay to show that Fc bound to immobilized FcRn can bind additional FcRn, which is only possible if there is more than one binding site for FcRn on Fc.

Isothermal titration calorimetry was used as the final verification that FcRn binds to IgG with a 2:1 stoichiometry. From these experiments, we calculated the thermodynamic parameters n , K_a and ΔH for FcRn titrating a human IgG1 monoclonal antibody (Carter *et al.*, 1992) and for IgG titrating FcRn. From the titration of FcRn with IgG, we obtained a value close to 0.5 Fc molecules bound per FcRn ($n = 0.38$), while the titration of IgG with FcRn yielded a value close to 2.0 FcRn molecules bound per Fc ($n = 2.05$). Both experiments yielded K_a values close to $2.0 \times 10^6 \text{ M}^{-1}$. This value suggests a lower affinity of receptor–IgG interaction than values obtained by Scatchard analysis of the binding of monoclonal rat IgG1 or IgG2b to detergent-isolated whole receptors purified from neonatal rat intestine ($K_a = 2 \times 10^7$ to $3 \times 10^7 \text{ M}^{-1}$; Hobbs *et al.*, 1987). The affinity difference may reflect a difference in method, a difference in the ligand (human IgG1 versus rat IgG2a or 2b), or a difference in the receptor (a truncated FcRn molecule lacking its transmembrane region in our study).

Prior to this report, the stoichiometry of purified FcRn binding to Fc had not been determined, but attempts were made to estimate the molecular

weight of the functional unit of IgG binding *in situ*. Electron irradiation of lyophilized and reconstituted intestinal brush borders suggested an approximate relative molecular mass of 110,000 (Simister & Rees, 1985), or 90,000 to 107,000 from preliminary studies on brush border membrane suspensions cooled to -100°C (Simister & Rees, 1983). These data suggested that a single FcRn heterodimer (relative molecular mass $\sim 65,000$) was not functional in IgG binding, and it was hypothesized that a disulfide linked dimer of heavy chains formed the functional receptor (Parham, 1989; Simister & Mostov, 1989a). However, we see no evidence of covalent dimerization of soluble FcRn by comparison of the mobility on SDS-PAGE gels run under reducing and non-reducing conditions, and purified FcRn elutes from a gel filtration column at the position expected for a complex of a single heavy and light chain (Gastinel *et al.*, 1992). In addition, the observation that the immobilized FcRn heterodimers still bind to Fc further suggests that a single FcRn heterodimer is functional in Fc binding. Our data suggest that two FcRn heterodimers can bind to a single Fc, but that the functional unit is not by necessity a dimer of two heavy chains with two $\beta 2$ -microglobulin subunits. However, juxtaposition of two FcRn heterodimers on intestinal membranes may facilitate IgG binding and transport, through binding of each Fc portion by two receptors. Knowledge of the FcRn-Fc stoichiometry of binding will facilitate the crystal structure determination of the complex, and serves as an initial characterization of the mode of FcRn interaction with its immunoglobulin ligand.

We thank M. Harrington for help with densitometry of gels and I. Tamir, D. C. Rees and M. Raghavan for critical reading of the manuscript. This work was supported by the Howard Hughes Medical Institute (P.J.B.). A.H.H. was supported by a pre-doctoral fellowship from the Howard Hughes Medical Institute. L.N.G. had a fellowship from the Centre National de la Recherche Scientifique during part of the work. P.J.B. is a Pew Scholar and has a young investigator Award from the Cancer Research Institute.

References

- Bjorkman, P. J. & Parham, P. (1990). Structure, function and diversity of class I major histocompatibility complex molecules. *Annu. Rev. Biochem.* **90**, 253–88.
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987a). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature (London)*, **329**, 512–518.
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987b). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature (London)*, **329**, 506–512.
- Carter, P., Presta, L., Gorman, C. M., Ridgway, J. B. B., Henner, D., Wong, W.-L. T., Rowland, A. M., Kottis, C., Carver, M. E. & Shepard, H. M. (1992). Humanization of an anti-P185HER2 antibody for human cancer-therapy. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 4285–4289.
- Chothia, C., Boswell, D. R. & Lesk, A. M. (1988). The outline structure of the T cell $\alpha\beta$ receptor. *EMBO J.* **7**, 3745–3755.
- Claverie, J.-M., Prochnicka-Chaloufour, A. & Bougueleret, L. (1989). Implications of a Fab-like structure for the T-cell receptor. *Immunology Today*, **10**, 10–14.
- Connolly, J. M., Potter, T. A., Wormstall, E. M. & Hansen, T. H. (1988). The Lyt-2 molecules recognizes residues in the class I $\alpha 3$ domain in allogeneic cytotoxic T-cell responses. *J. Exp. Med.* **168**, 325–341.
- Culbertson, J. T. (1938). Natural transmission of immunity against *Trypanosoma lewisi* from mother rats to their offspring. *J. Parasitol.* **24**, 65–82.
- Davis, M. M. & Bjorkman, P. J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature (London)*, **334**, 395–402.
- Deisenhofer, J. (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. *Biochemistry*, **20**, 2361–2370.
- Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A. & Wilson, I. A. (1992). Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science*, **257**, 919–927.
- Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & D. C. Wiley, (1989). Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature (London)*, **342**, 692–696.
- Gastinel, L. N., Simister, N. E. & Bjorkman, P. J. (1992). Expression and crystallization of a soluble and functional form of an Fc receptor related to class I histocompatibility molecules. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 638–642.
- Hobbs, S. M., Jackson, L. E. & Peppard, J. V. (1987). Binding of subclasses of rat immunoglobulin G to detergent-isolated Fc receptor from neonatal rat intestine. *J. Biol. Chem.* **262**, 8041–8046.
- Hope, H. (1990). Crystallography of biological macromolecules at ultra-low temperature. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 107–126.
- Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L. & Ohlendorf, D. H. (1987). The use of an imaging proportional counter in macromolecular crystallography. *J. Appl. Crystallogr.* **20**, 383–387.
- Jancarik, J. & Kim, S. H. (1991). Sparse-matrix sampling—a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24**, 409–411.
- Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. (1991). The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature (London)*, **353**, 321–325.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
- McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*. John Wiley and Sons, New York.
- Parham, P. (1989). MHC meets mother's milk. *Nature (London)*, **337**, 118–119.
- Rodewald, R. (1976). pH-dependent binding of immunoglobulins to intestinal cells of the neonatal rat. *J. Cell Biol.* **71**, 666–670.
- Salter, R. D., Norment, A. M., Chen, B., Clayberger, C., Krensky, A. M., Littman, D. & Parham, P. (1989). A binding site for the T-cell co-receptor CD8 on the $\alpha 3$ domain of HLA-A2. *Nature (London)*, **338**, 345–347.
- Simister, N. E. & Mostov, K. E. (1989a). Cloning and expression of the neonatal rat intestinal Fc receptor,

- a major histocompatibility complex class I antigen homolog. *Cold Spring Harbor Symp. Quant. Biol.* **53**, 571–580.
- Simister, N. E. & Mostov, K. E. (1989b). An Fc receptor structurally related to MHC class I antigens. *Nature* **337**, 184–187.
- Simister, N. E. & Rees, A. R. (1983). *Properties of Immunoglobulin G-Fc Receptors from Neonatal Rat Intestinal Brush Borders*. Ciba Foundation symposium No. 95, Pitman Books Ltd., London.
- Simister, N. E. & Rees, A. R. (1985). Isolation and characterization of an Fc receptor from neonatal rat small intestine. *Eur. J. Immunol.* **15**, 733–738.
- Townsend, A. & Bodmer, H. (1989). Antigen recognition by class I-restricted T lymphocytes. *Annu. Rev. Immunol.* **7**, 601–624.
- Wiseman, T., Williston, S., Brandts, J. F. & Lin, L.-N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137.

Edited by A. Klug

Appendix B

The Relative Affinity of Monoclonal and Polyclonal IgG for a Rat Intestinal Fc Receptor

Introduction

FcRn, an Fc receptor transiently expressed by the intestinal epithelia of neonatal rats, functions to transport maternal immunoglobulin G (IgG) from the lumen of the intestine to the bloodstream. FcRn was predicted to be structurally similar to class I major histocompatibility complex (class I MHC) molecules based on an analysis of primary structure¹. Despite the apparent structural similarity, FcRn and class I MHC molecules bind very different physiological ligands. FcRn binds the Fc portion of whole IgG (~ 160 kDa) while class I MHC molecules bind peptides of approximately 9 amino acids (~ 1 kDa). The manner in which FcRn interacts with Fc is of considerable interest because it may provide insight into how class I MHC molecules interact with members of the immunoglobulin gene superfamily, such as T-cell receptors and CD8. The structure of an FcRn-IgG complex is also of general interest because it would reveal how the same structural framework can be utilized for very different biological functions.

The crystallographic structure of an FcRn-Fc complex was pursued because the three-dimensional atomic structure of such a complex would permit the function of FcRn to be understood at a molecular level (Appendix A). Crystals were initially obtained using a purified, soluble form of FcRn² and the Fc portion of polyclonal rat IgG. Unfortunately, these crystals diffract to relatively low resolution (5.5Å, with anisotropic diffraction to 3.5Å in the direction of the *a*-axis) and are highly sensitive to radiation damage. It was proposed that flash cooling the crystals to N₂(l) temperatures might allow the collection of high resolution data that could not be observed at room temperature due to radiation damage. Flash freezing did eliminate radiation decay but no improvements were observed in the diffraction limits of the crystals (A. Huber, unpublished results). In order to facilitate the solution of a low resolution FcRn-Fc crystal structure and to better characterize FcRn function, the stoichiometry of the FcRn-IgG complex was determined biochemically

(Appendix A). This work demonstrated that two molecules of FcRn can bind to a single IgG molecule and that the stoichiometry of binding in the FcRn-Fc crystals is two FcRn per Fc.

While crystallography does yield tremendous insight into molecular structure and function, it can only provide a static picture of dynamic molecules and processes. Thus, crystallographic structure information generally needs to be complemented with a biochemical analysis of function. Biochemical characterization is sometimes a matter of necessity, especially when a crystallographic analysis has stalled. Since it was not clear that the poorly diffracting complex crystals could be solved, a study of the relative affinity of FcRn for different IgG molecules was undertaken. This type of analysis can identify high affinity complexes for crystallization and could lead to the localization of the ligand binding site. Information concerning the location of the FcRn binding site on IgG would facilitate the solution of a low resolution complex structure.

Competition assays are frequently used to measure the relative affinities of ligands for a common receptor³. The relative affinities of different IgG for cell surface FcRn can be easily measured using a competitive radioimmunoassay (Louis N. Gastinel, personal communication). This assay measures the ability of unlabeled (sample) IgG to inhibit the binding of radiolabeled (standard) IgG by lipid-linked FcRn expressed on the surface of Chinese hamster ovary (CHO) cells². Measurements are taken at several different concentrations of sample IgG and are used to calculate the concentration of sample IgG required to inhibit binding of the standard by 50 percent (I_{50}).

If two monoclonal immunoglobulins are shown to have very different affinities for an Fc receptor, chimeric molecules can be used to determine which part of the immunoglobulin is being recognized by the receptor. A large difference in the affinities of human IgG3 and IgG2 for Fc γ RI, a high affinity IgG Fc-receptor, was used to localize the

region of IgG bound by the receptor⁴. Chimeric IgG molecules were produced by swapping domains between human IgG3 and IgG2 and the relative affinities of the chimeric molecules for Fc γ RI was determined. These experiments demonstrated that the subclass identity of the immunoglobulin C_H2 domains modulates Fc γ RI binding. Chimeric immunoglobulins could potentially be used to identify where FcRn binds IgG. This knowledge would significantly advance our understanding of FcRn function and would facilitate the generation of physiologically relevant Fc fragments that might yield high quality FcRn-Fc co-crystals.

An examination of FcRn's affinity for different immunoglobulins might identify promising candidates for chimera studies and could also lead to the localization of the FcRn binding site through sequence comparisons. Differences in affinity should be reflected in heavy chain constant domain sequence differences. If a sufficient number of immunoglobulins were tested, one might find sequences that are conserved only within immunoglobulins that have a high affinity for FcRn. This type of analysis was used to identify regions within the IgG C_H2 domain that were likely to interact with Fc γ RI⁴. As mentioned earlier, such a localization of the FcRn binding site would not only better characterize FcRn function, but might lead to crystallization of physiologically relevant complexes of FcRn and a fragment of Fc. Identifying which IgGs have a high affinity for FcRn might also facilitate the growth of better FcRn-Fc or FcRn-IgG cocrystals. Complexes produced with high affinity IgG or Fc ligands might be more stable than the FcRn-Fc complex used in the original crystallization trials. The additional stability of the complex and differences in Fc sequence could lead to the growth of new crystal forms that diffract to higher resolution.

Here I report the results of competitive binding assays that measured the affinity of FcRn for different human IgG subclasses as well as polyclonal IgGs from a variety of animal species. The experiments with human IgG subclasses were done as a preliminary

test to see if it was feasible to use human chimeric IgG proteins⁴ to localize the FcRn IgG binding site. The relative affinities of bovine, goat, rabbit, human and rat polyclonal IgG for FcRn were also measured. These experiments were carried out in order to determine which IgG should be used in future complex crystallization trials. The results of the polyclonal affinity assays could not be used for a sequence-based localization of the FcRn binding site because the polyclonal IgG contain a variety of IgG subclasses.

Materials and Methods

Reagents

Dialyzed fetal bovine serum and fetal bovine serum were from GIBCO/BRL. Methionine sulfoximine was from Sigma. α -minimum essential medium was from Irvine Scientific. AFRC-MAC 193, a rat IgG2a, was a gift from Geoffrey Butcher (AFRC, Babraham Institute) and ¹²⁵I labeled AFRC-MAC 193 was prepared by Louis N. Gastinel.

Cell Suspensions

CHO cells expressing FcRn with a lipid-linked heavy chain (p51-DAF)² were cultured in α -minimum essential medium supplemented with 10% dialyzed fetal bovine serum, 100 units/ml penicillin/streptomycin and 500 μ M methionine sulfoximine. Wild-type CHO cells were grown in α -minimum essential medium with 10% dialyzed fetal bovine serum and 100 units/ml penicillin/streptomycin.

Adherent p51-DAF expressing cells and untransfected CHO cells were removed from culture dishes by incubation with phosphate buffered saline (PBS), 10 mM EDTA for 10 minutes at 37° C. Pelleted cells were resuspended in 50 mM Phosphate-NaOH, pH 6, 150 mM NaCl, 1% (w/v) BSA, to a final concentration of 2x10⁶ cells/ml.

Binding Assay

Each competitive inhibition assay contained: ^{125}I labeled rat IgG2a (AFRC-MAC 193), competitor cold immunoglobulin and approximately 2×10^5 p51-DAF cells. Assays were brought to a final volume of 500 μL with 50mM phosphate, 150 mM NaCl, 1% (w/v) BSA, pH 6. Assays containing no competitor IgG were used to determine the signal for 0% inhibition and the level of background was estimated by using wild-type CHO cells and no competitor.

After incubation for 1 hour at room temperature, the assays were centrifuged in a model 5415 Eppendorf Micro Centrifuge for 1 minute at 14,000 rpm. Supernatants were discarded and the pellets were washed with 1 mL of cold 50 mM phosphate, 150 mM NaCl, pH 6. The samples were centrifuged a second time and the supernatants were discarded. The pellets were counted with a gamma counter.

Calculation of I_{50}

The signal for 0% inhibition (Total) was calculated by subtracting the estimated background from the number of counts observed without competitor. The number of counts for each sample (Sample) was also adjusted by subtracting the background. The percent inhibition in a given assay was calculated as follows:

$$\% \text{ inhibition} = [(\text{Total} - \text{Sample}) / \text{Total}] \times 100$$

The concentration of competitor required to inhibit binding of the labeled rat IgG2a by 50 percent (I_{50}) was estimated by plotting percent inhibition versus $\log(\text{competitor concentration, nM})$ for each inhibitor IgG. When plotted in this manner, the inhibition curve is commonly sigmoidal with a linear range from approximately 20-80% inhibition. A

linear line of best fit was calculated for each competitor using data points from competitor concentrations that resulted in 20-80% inhibition. This line was used to estimate I_{50} .

Results and Discussion

Affinity of FcRn for Human IgG Isotypes

The concentrations of human IgG1, IgG2, IgG3 and IgG4 estimated to cause 50% inhibition of binding were all within the same order of magnitude (Fig. B-1A-E and Table B-1). Thus, all four subclasses of human IgG were bound by FcRn with similar affinities, with IgG1 and IgG2 having the highest affinity and human IgG4 having the lowest (Fig. B-1A-D and Table B-1). Since no substantial differences in affinity were observed, it was not feasible to use human IgG chimeric proteins⁴ to identify the region of Fc bound by FcRn.

Affinity of FcRn for Polyclonal IgG from Different Animal Species

The average affinities of polyclonal bovine, goat, human, rabbit, and rat IgG for FcRn were estimated by competitive radioimmunoassay. FcRn has a very low affinity for bovine IgG with less than 15% inhibition observed at a bovine IgG concentration of 320 nM (data not shown). The relative affinities of the other IgG for FcRn, as determined by estimated I_{50} values, are as follows: rabbit > human > rat >> goat >> bovine (Table B-1). Given the strong binding observed for rabbit and human IgG, it is likely that the Fc portions of these molecules bind FcRn with a higher affinity than rat Fc. An increase in complex stability coupled with changes in the amino acid sequence of the Fc could lead to new and better FcRn-Fc cocrystals. However, preliminary crystallization trials with rabbit and human IgG have not produced crystals that diffract to a higher resolution than those previously reported.

Table B-1 Estimated concentration of IgG required to inhibit FcRn binding of an ^{125}I labeled rat IgG2a by 50 percent (I_{50}).

<u>Immunoglobulin Tested</u>	<u>I_{50}, nM</u>
Human IgG1	20
Human IgG2	20
Human IgG3	35
Human IgG4	70
Rat IgG2a	50
Goat polyclonal IgG	400
Human polyclonal IgG	30
Rabbit polyclonal IgG	10
Rat polyclonal IgG	55
Bovine polyclonal IgG	>1,000

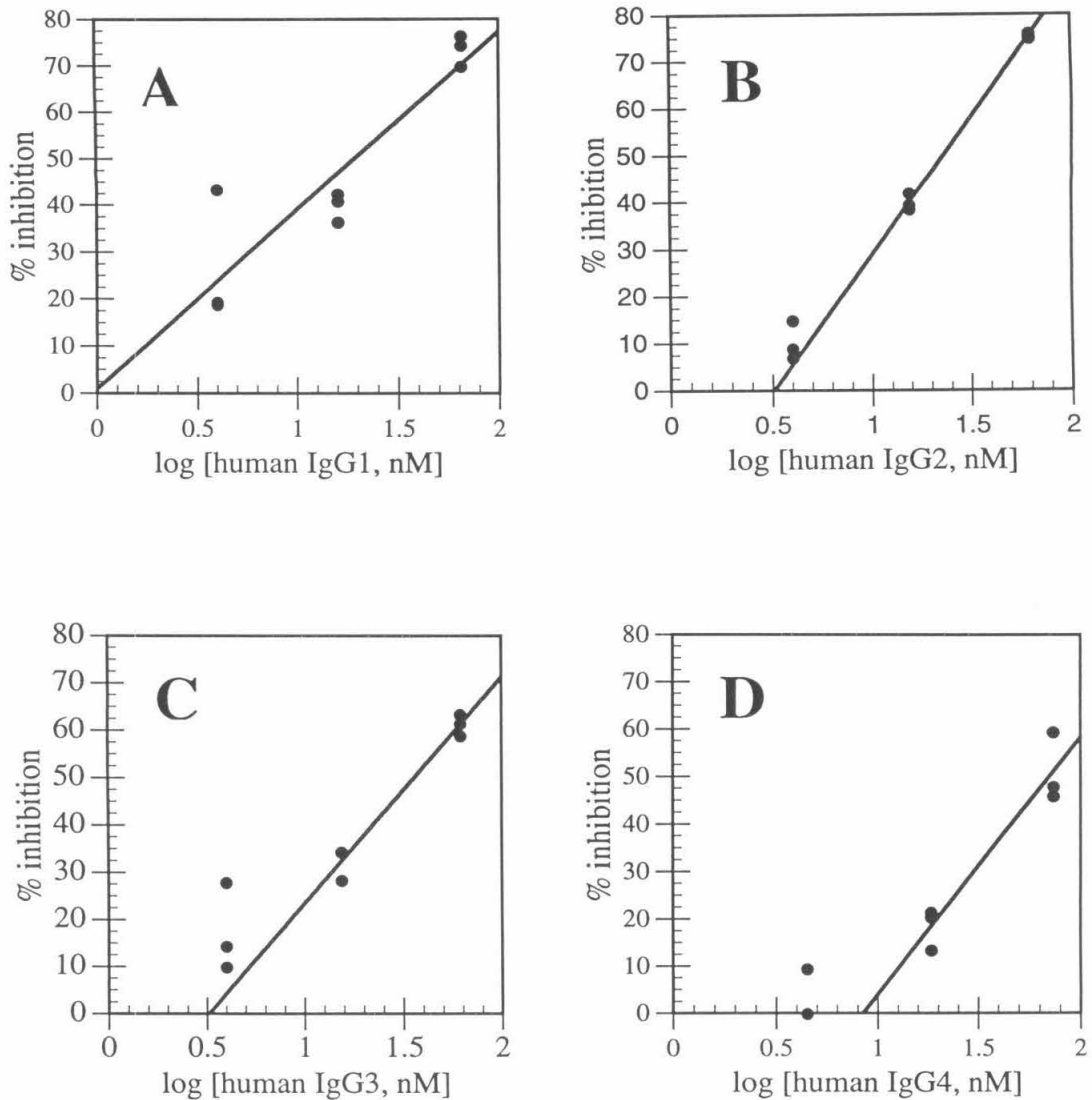
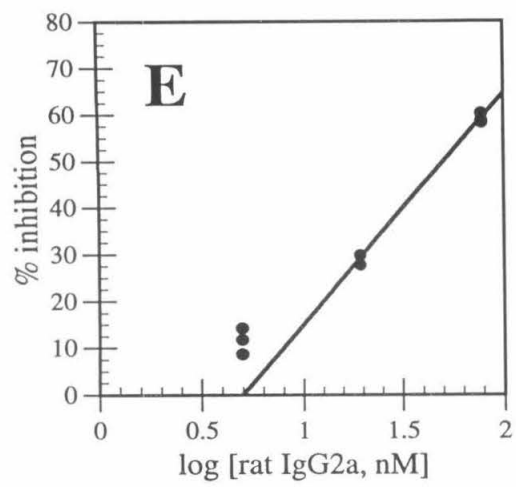


Figure B-1 Results of competitive inhibition assays for purified subclasses of IgG. Percent inhibition is plotted as a function of log [competitor, nM] for purified human IgG1 (A), IgG2 (B), IgG3 (C), IgG4 (D) and rat IgG2a (E). Data points for IgG concentrations yielding 20-80% inhibition were used to plot the line of best fit. Three replicates were measured for each IgG concentration.



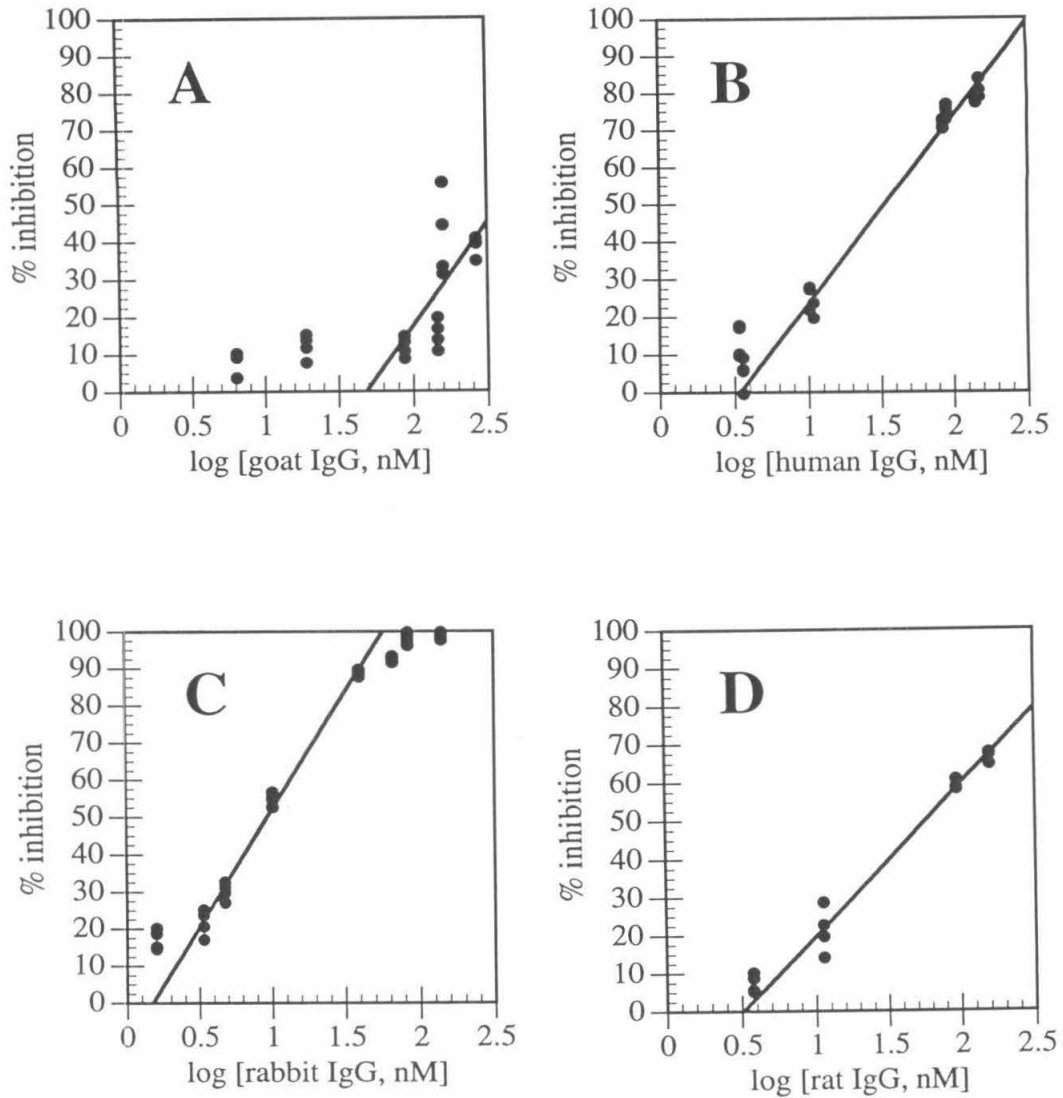


Figure B-2 Results of polyclonal IgG competitive inhibition assays. Percent inhibition is plotted as a function of log [competitor, nM] for goat (A), human (B), rabbit (C) and rat (D) polyclonal IgG. Data points for IgG concentrations yielding 20-80% inhibition were used to plot the line of best fit. Four replicates were measured for each IgG concentration.

References

1. Simister, N.E. & Mostov, K.E. (1989). An Fc receptor structurally related to MHC class I antigens. *Nature* **337**, 184-187.
2. Gastinel, L.N., Simister, N.E. & Bjorkman, P.J. (1992). Expression and crystallization of a soluble and functional form of an Fc receptor related to class I histocompatibility molecules. *Proc. Natl. Acad. Sci USA* **89**, 638-642.
3. Müller, R. (1983). Determination of affinity and specificity of anti-hapten antibodies by competitive radioimmunoassay. *Meth. Enzymol.* **92**, 589-601.
4. Canfield, S.M. & Morrison, S.L. (1991). The binding affinity of human IgG for its high affinity Fc receptor is determined by multiple amino acids in the CH2 domain and is modulated by the hinge region. *J. Exp. Med.* **173**, 1483-1491.